Winter 7-19-2019

An Assessment of the National Water Model's Ability to Reproduce Drought Series in New York State

Brenden Covert
covertb9@yahoo.com

AN ASSESSMENT OF THE NATIONAL WATER MODEL'S ABILITY TO REPRODUCE DROUGHT SERIES IN NEW YORK STATE

by

Brenden Covert

a thesis
submitted in partial fulfillment
of the requirements for the
Master of Science Degree
State University of New York
College of Environmental Science and Forestry
Syracuse, New York
July 2019

Graduate Program in Environmental Resources Engineering

Approved by:
Charles N. Kroll, Major Professor
Jonathan Cohen, Chair, Examining Committee
Lindi Quackenbush, Department Chair
S. Scott Shannon, Dean, The Graduate School

Acknowledgements

I would like to thank my family and friends for their continued support throughout this process. This would not have been possible without the consistent encouragement from my parents, brother, and my friends who always helped me keep a good head on my shoulders. I would also like to thank all of my grandparents who, in one way or another, invested heavily in the successes I was able to achieve over the past 6 years in both my undergraduate and masters work.

I would like to thank SUNY ESF, and more specifically the ERE department for the financial support throughout the duration of my master's degree. ESF, and its community, will forever hold a special place in my heart for the integral role the 6 years I have spent on campus have played in my development personally, academically, and as an environmental steward. I would like to thank my professors for the positive encouragement that helped me push past the many obstacles encountered in both my undergrad and graduate studies at ESF.

Finally, I would like to thank Chuck for affording me this opportunity to go attend graduate school. Without him, I could not have gotten to where I am today. I can't imagine a better mentor I've had the pleasure of working with over the past several years. I always felt like it was impossible to be stressed under his direction, which made completion of this masters a pretty enjoyable experience.

Table of Contents

List of Tables

List of Figures

Abstract

B.H. Covert. An Assessment of the National Water Model's Ability to Reproduce Drought Series in New York State. 126 Pages, 4 Tables, 18 Figures, 2019. APA Style Guide used.

Accurately quantifying the spatiotemporal extent and intensity of drought is important for water resources planning and management. The onset of drought is impacted by a multitude of meteorologic, climatologic, and anthropogenic triggers. Currently, the United States Drought Monitor (USDM) is the tool used in the United States (U.S.) to both categorize drought and allocate emergency funding based on drought categories. Two key issues with the USDM are the nonuniform scale of its input variables, and its inability to predict drought into the future. This work proposes employing output from the NOAA's National Water Model (NWM) 25-year reanalysis to augment the USDM. A numerical comparison is presented between concurrent New York State Mesonet (Mesonet) observed and NWM simulated soil moisture data at three depths in the soil column for 119 locations across NY, and between USGS observed and NWM simulated streamflow data for 28 locations across NY. Drought categories are determined according to USDM percentile ranges based on the percentiles derived from the non-exceedance probabilities of NWM streamflow and soil moisture output and USGS streamflow data. It was determined that while there was mixed (streamflow) to poor (soil moisture) agreement between the NWM reanalysis and observational data, the NWM was generally able to reproduce the spatial extent of the 2016 drought in NY with reasonable skill, especially when using data representing deeper soil moisture. It was generally observed that NWM derived drought categories were typically more extreme than those of the USDM.

Key words: drought, streamflow, soil moisture, New York State Mesonet, National Water Model

B.H. Covert
Candidate for the degree of Mater of Science
Charles N. Kroll, Ph.D
Department of Environmental Resources Engineering
State University of New York College of Environmental Sciences and Forestry
Syracuse New York

Chapter 1: Introduction

1.1 Problem Overview and Motivation

The United Nations (UN), as a part of its Envision Disability 2030 campaign, established 17 Sustainable Development Goals to transform the world. Goal 6 reflects on Clean Water and Sanitation (UN 2019a). The targets of Goal 6 include implementing integrated water resources management by increasing water use efficiency to address water scarcity and reduce the number of those affected by it (UN 2019b). The Centers for Disease Control and Prevention Global Water, Sanitation and Hygiene department (CDC WASH) reports that currently there are 780 million people globally who do not have access to a safe and sustainable drinking water supplies (CDC 2019). With increasing atmospheric temperatures, there is expected to be a general increase in potential evapotranspiration (ET), which is likely to further exacerbate drought conditions and water demand (Cook et al. 2015, Fischer et al. 2007, IPCC 2007, Held & Soden 2006). As the world's population continues to rise, the need to effectively manage, understand and utilize water resources is of paramount concern (Winz et al. 2009), especially during periods of drought.

Common to all types of drought is a deficiency of precipitation that results in a shortage of water (Wilhite & Glantz 1985). In general, drought tends to decrease soil moisture, streamflow, and surface and groundwater supplies (Hayohe et al. 2007, NRCC 2017). There are generally four primary ways to characterize a drought: meteorologic, hydrologic, agricultural, and societal. A meteorologic drought is instigated by a deficiency in precipitation over an extended period of time. Many of the other types of drought emanate from a meteorological drought (Hao & Singh 2015). An agricultural drought arises when there is insufficient water available for plant growth or crop production (Hao & Singh 2015); this "short-term" drought is often associated with a decline to soil moisture. Under a sustained lack of precipitation, streamflow, groundwater and reservoir levels may decline, resulting in hydrologic drought; this is generally considered a "long-term" drought (Hao & Singh 2015). A societal drought is often associated

with the disruption of the supply and demand of economic goods and results from the three aforementioned types of droughts (Hao & Singh 2015). All of these types of drought have ecological, economic, and societal ramifications (Wilhite 2000).

Recently, the United States (U.S.) has had several costly drought events, particularly in Georgia (2007-2012), Texas (2010-2013) and California (2012-1014) (Rippey 2015, Kohl & Knox 2016, Flanders et al. 2008, Nielsen-Gammon 2011, Guerrero 2012, Howitt et al. 2014). The Northeastern U.S., which is generally considered to be a "water rich" region, was also recently affected by a drought (Sweet et al. 2017). As droughts are becoming both more common in manifestation and extreme in nature, the ability to accurately characterize their extent and intensity is important (Hayohe et al. 2007). To that end, this analysis is focused on the characterization of drought in New York (NY), with emphasis placed on the spatial and temporal extent of the 2016 drought.

The use of a modeling platform to explore and understand the spatiotemporal nature of drought has been relatively unexplored. Since droughts are instigated by a multitude of climatologic, hydrologic and anthropogenic triggers, understanding their evolution and diversity is challenging, especially when attempting to model their occurrence (Mishra et al. 2015). Modeling of droughts has often been statistical in nature and focused on reproducing one or more drought indices (e.g. the Standardized Precipitation Index (SPI)). For example, Mirsha & Desi (2005) employed a stochastic process to predict the SPI for up to a 2-month lead time and found that the mean and variance of the observed and the modeled drought index were not significantly different. However, the application of deterministic spatiotemporal models to predict drought has been somewhat limited (Mishra & Singh 2011).

One prominent example of an attempt at modeling the spatiotemporal extent of drought was conducted by Xia et al. (2012, 2014). Xia et al. (2012) sought to compare streamflow output from the North American Land Data Assimilation System Phase 2 (NLDAS-2) to USGS gauged streamflow data at 961 small

basins across the continental United States (CONUS). In general, a Land Data Assimilation System integrates satellite and ground-based observational data products, land surface modeling, and data assimilation techniques to produce high quality estimators of land surface states (e.g., soil moisture, temperature) and environmental fluxes (e.g., ET, surface runoff) (NASA 2019) One objective of NLDAS-2 is to support the National Drought Information System (NIDIS), which has the general goal of facilitating drought research through a collaborative network consisting of federal, tribal, state and local partnerships (Xia et al. 2010, NIDIS 2019). Xia et al. (2012) generally concluded that multimodel products, specifically NLDAS-2, are well suited for drought classification. With that encouraging finding, Xia et al. (2014) aimed to assess the ability of NLDAS-2 to reproduce drought categories for six large U.S. Drought Monitor (USDM) regions (West, South, High Planes, Midwest, Southeast, Northeast), and for all states across the CONUS. They used NLDAS-2 streamflow, soil moisture, and ET to develop a blended drought index, and then assessed the ability of this index to reproduce USDM classifications. They found that NLDAS-2 soil moisture in the top 1 meter of the soil column and the total soil moisture appears to have more skill in predicting USDM classifications than NLDAS-2 streamflow and ET, concluding that soil moisture plays an integral role in accurately categorizing drought (Xia et al. 2014).

For this thesis, the use of output from the National Water Model (NWM), which has a similar structure to NLDAS-2, to characterize the spatiotemporal distribution of drought series in NY will be explored. The NWM, a deterministic hydrologic modeling framework, is developed collaboratively by consortium of Universities, the NOAA National Weather Service (NWS) and the USGS (Sullivan 2016). The NWM has had past success in modeling hydrologic extremes, particularly in modeling floods (Lamont et al. 2018). Like droughts, floods are triggered by a host of environmental influences and have been reasonably successfully modeled since the early 1970s (Goel et al. 1998). Lamont et al. (2018) recently employed the NWM to estimate flood frequency values at each of the approximately 2.7 million stream links in the National Hydrography Dataset (NHD). When these values were converted to inundation maps (at a 10 m horizontal resolution), promising agreement was shown with the current Federal Emergency Management

Agency (FEMA) maps for the 100-Year floodplain. The NWM runs on a relatively fine horizontal resolution (generally ~1 km) and requires a similar set of input variables to that of NLDAS-2 (e.g., high-resolution radar-gauge precipitation inputs, numerical weather predictions) (NOAA 2019, Xia et al. 2012).

Due to its past success at modeling hydrologic extremes, and its similarity in form to NLDAS-2, the NWM is proposed as a candidate to augment information sources that comprise the USDM. The USDM is one of the primary tools in the U.S. to both characterize the current state of drought and allocate federal funding for drought relief aid (USDA 2017). There are several known issues with the USDM. The USDM employs data on a non-uniform scale, uses inconsistent data products to produce maps in specific regions, and only provides a determination of the current spatial extent and severity of drought. The USDM currently has no ability to predict drought in the future. Use of the NWM may help to provide more homogeneity in scale to the USDM and should the NWM show skill in characterizing drought similarly to the USDM, it may provide a platform to allow the forecasting of future drought events which would be critically important for water resources planning and management.

Access to accurate localized data, particularly that of soil moisture, is important for characterizing the extent and intensity of drought (Sheffield & Wood 2007). NY is home to a relatively new network of 126 meteorological stations termed the NYS Mesonet (Mesonet). The Mesonet, which was developed to provide an advanced early warning weather-detection system, provides a spatially rich set of meteorological data (including soil moisture) through a network of technologically advanced sensors (NYSM 2019). Given the availability of Mesonet data and NWM reanalysis output, NY is an excellent study area to explore how modeling and data applications may enhance the USDM. Here we propose to examine: (1) how NWM streamflow output compares to USGS unregulated gauged streamflow data in NY, (2) how NWM soil moisture output compares to soil moisture measured at Mesonet sites and at a limited number of longer-term soil moisture stations in NY, and (3) how drought intensity predicted by

the NWM at streamflow and soil moisture sites compares to drought intensity characterized by the USDM, with a specific focus on the drought of 2016.

1.2 Research Goals

The goals of this study are to explore methods to improve the characterization of drought in New York. National Water Model (NWM) streamflow and soil moisture output will be analyzed and compared to gauged streamflow and measured soil moisture series. Use of NWM output to reproduce the USDM will also be explored. More specifically, this study aims to:

1) Examine how NWM simulated streamflow and soil moisture output compares to observed streamflow and soil moisture during drought periods.
2) Explore how the NWM characterizes the intensity of drought in NY and to compare this to drought intensity from the USDM, with a particular focus on the Northeast drought of 2016.

1.3 Organization of Chapters

This thesis is presented in three chapters. Chapter 1 (Introduction) is an introductory chapter which presents the motivation for this research and our research goals. Chapter 2 (National Water Model's Ability to Reproduce Drought Series) presents the main analysis and specifically aims to address the two aforementioned research goals. Chapter 2 is set up like a stand-alone journal article, with sections including an introduction, data and methods, results and discussion, and conclusions. Chapter 3 (Thesis Summary and Next Steps) reflects back on the results from Chapter 2 and how those results address the research goals presented in Chapter 1. In addition, Chapter 3 contains a discussion of the limitations and potential next steps of this analysis. Finally, Appendix 1, which is also set up like a stand-alone journal article, contains another unrelated research experiment, An Assessment of i-Tree Hydro, where a sensitivity analysis and various model calibration techniques are implemented on a series urban catchments in Baltimore, MD to assess the performance of i-Tree Hydro.

1.4 References

Centers for Disease Control and Prevention Global Water Sanitation and Hygiene (WASH). (2019). Global WASH Fast Facts. Retrieved on June 30, 2019, from https://www.cdc.gov/healthywater/global/wash_statistics.html.

Cook, B. I., Ault, T. R., & Smerdon, J. E. (2015). Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Science Advances*, *1*(1), e1400082.

Fischer, G., Tubiello, F. N., Van Velthuizen, H., & Wiberg, D. A. (2007). Climate change impacts on irrigation water requirements: effects of mitigation, 1990–2080. *Technological Forecasting and Social Change*, *74*(7), 1083-1107.

Flanders, A., E. Bauske, and J. McKissick, 2008: Economic impact of total watering restrictions to the green industry in the Bear Creek Reservoir region. Center for Agribusiness and Eco- nomic Development Rep. CR-08-05, University of Georgia, 8 pp. [Available online at http://www.caes.uga.edu/center/caed/ pubs/2008/documents/CR-08-05.pdf.]

Goel, N. K., Seth, S. M., & Chandra, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, *124*(2), 146-155.

Guerrero, B. (2012). The impact of agricultural drought losses on the Texas economy, 2011. *Briefing Paper, AgriLife Extension*.

Hao, Z., & Singh, V. P. (2015). Drought characterization from a multivariate perspective: A review. *Journal of Hydrology*, *527*, 668-678.

Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J., DeGaetano, A., Troy, T.J., & Wolfe, D. (2007). Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dynamics*, *28*(4), 381-407.

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of climate*, *19*(21), 5686-5699.

Howitt, R., Medellín-Azuara, J., MacEwan, D., Lund, J. R., & Sumner, D. (2014). *Economic analysis of the 2014 drought for California agriculture*. University of California, Davis, CA: Center for Watershed Sciences.

Intergovernmental Panel on Climate Change (IPCC). (2007). Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K., Reisinger, A. (Eds.)]. IPCC, Geneva, Switzerland, 10 pp.

Kohl, E., & Knox, J. A. (2016). My drought is different from your drought: A case study of the policy implications of multiple ways of knowing drought. *Weather, Climate, and Society*, *8*(4), 373-388.

Lamont, S., Munchak, L., & Gray, L. (2018). A Continental Scale Fluvial Flood Risk Assessment Based on the NOAA National Water Model Reanalysis Dataset. In *AGU Fall Meeting Abstracts*, *December 2018.*

Mishra, A. K., & Desai, V. R. (2005). Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment*, *19*(5), 326-339.

Mishra, A. K., & Singh, V. P. (2011). Drought modeling–A review. *Journal of Hydrology*, *403*(1-2), 157-175.

Mishra, A. K., Sivakumar, B., & Singh, V. P. (2015). Drought processes, modeling, and mitigation. *Journal of Hydrology*, *526*, 1-2.

National Atmospheric and Space Administratin (NASA). (2019). Land Data Assimilation System, LDAS. Retrieved August 10, 2019 from https://ldas.gsfc.nasa.gov/ldas/land-data-assimilation-system

National Integrated Drought Information System (NIDIS). (2018). What is NIDIS? Retrieved on July 2, 2018, from https://www.drought.gov/drought/what-nidis.

National Oceanic and Atmospheric Administration Office of Water Protection. (2019). Information About the National Water Model. Retrieved June 20, 2019, from http://water.noaa.gov/about/nwm.

New York State Mesonet. (2019). About New York's Mesoscale Weather Network. Retrieved June 30, 2019, from http://www.nysmesonet.org/#about.

Nielsen-Gammon, J. (2011). *The 2011 Texas Drought: A Briefing Packet for the Texas Legislature, October 31, 2011,* The Office of the State Climatologist, Texas.

Northeast Regional Climate Center (NRCC), Cornell University (2016). Coping with Drought and its Aftermath in the Northeast. Retrieved on March 4, 2019, from http://www.nrcc.cornell.edu/regional/drought/pubs/assessment_2016.pdf.

Rippey, B. R. (2015). The US drought of 2012. *Weather and Climate Extremes*, *10*, 57-64.

Sheffield, J., & Wood, E. F. (2007). Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *Journal of Geophysical Research: Atmospheres*, *112*(D17).

Sullivan, K. (2016). *National Water Model Improving NOAA's Water Prediction Services* (United States, National Oceanic and Atmospheric Administration). Retrieved January 22, 2019, from http://water.noaa.gov/documents/wrn-national-water-model.pdf.

Sweet, S. K., Wolfe, D. W., DeGaetano, A., & Benner, R. (2017). Anatomy of the 2016 drought in the Northeastern United States: implications for agriculture and water resources in humid climates. *Agricultural and Forest Meteorology*, *247*, 571-581.

United Nations Department of Economic and Social Affairs. (2019a). Envision 2030: 17 Goals to Transform the World for Persons with Disabilities. Retrieved on June 30, 2019, from https://www.un.org/development/desa/disabilities/envision2030.html.

United Nations Department of Economic and Social Affairs. (2019b). Envision 2030 Goal 6: Clean Water and Sanitation. Retrieved on June 30, 2019, from https://www.un.org/development/desa/disabilities/envision2030-goal6.html.

United States Department of Agriculture Farm Service Agency. (2019). Disaster Assistance Fact Sheet, October 2017: Emergency Disaster Designations and Declaration Process. Retrieved on July 3, 2019, from https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/FactSheets/2017/emergency_disaster_designation_and_declaration_process_oct2017.pdf.

Wilhite, D. A. (2000). Chapter 1 Drought as a Natural Hazard: Concepts and Definitions. (2000). *Drought Mitigation Center Faculty Publications,* 69, http://digitalcommons.unl.edu/droughtfacpub/69

Wilhite, D. A., & Glantz, M. H. (1985). Understanding: the drought phenomenon: the role of definitions. *Water international*, *10*(3), 111-120.

Winz, I., Brierley, G., & Trowsdale, S. (2009). The use of system dynamics simulation in water resources management. *Water resources management*, *23*(7), 1301-1323.

Xia, Y., Ek, M. B., Peters-Lidard, C. D., Mocko, D., Svoboda, M., Sheffield, J., & Wood, E. F. (2014). Application of USDM statistics in NLDAS-2: Optimal blended NLDAS drought index over the continental United States. *Journal of Geophysical Research: Atmospheres*, *119*(6), 2947-2965.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y. Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

Chapter 2: National Water Model's Ability to Reproduce Drought Series

2.1 Introduction

Water is a critical resource necessary to sustain life (Shiklomanov 1998). Lack of water can lead to ecological degradation, stress on agricultural and hydropower production, and threats to human and ecosystem health (Postel et al. 1996). Despite efforts to reduce water consumption (Donnelly & Cooley 2015), there are often periodic drought events throughout the world, and the UN predicts that by 2025, 1.8 billion people will experience water scarcity and two-thirds of the world's population will live under water-stressed conditions (Spinoni et al. 2014, UN 2009). In general, drought tends to decrease soil moisture, streamflow, and supplies of surface and groundwater (Hayohe et al. 2007, NRCC 2017). There is also evidence that a changing climate, population and economic growth, and changes to land use and energy generation are placing further stress on our limited water systems. Consequently droughts are becoming both more common in manifestation and extreme in nature (Bradford 2018, Hayohe et al. 2007).

Droughts can be economically catastrophic, as demonstrated by several recent events in the United States (U.S.), such as the droughts in Georgia (2007-2012), Texas (2010-2013) and California (2012-2014), where total economic losses were estimated to be in the billions of dollars (Rippey 2015, Kohl & Knox 2016, Flanders et al. 2008, Nielsen-Gammon 2011, Guerrero 2012, Howitt et al. 2014). The Northeastern U.S., which is generally considered a "water rich" region, has also been affected by recent drought events. The summer of 2016 was one of the warmest and driest on record (Sweet et al. 2017). Compounded by historically low snowfall in the preceding winter, the Northeast experienced one of the worst droughts since the 1960s (which is generally considered to be the drought of record in the Northeastern U.S.) (Seager et al. 2012). In New York (NY) both rainfall and streamflow declined substantially from historical averages for most of the 2016 growing season (Sweet et al. 2017). Topsoil was the driest it had been in over 20 years, and hindered by a lack of established irrigation infrastructure, many farmers

reported greater than 30% losses to their rainfed crops (Sweet et al. 2017, NRCC 2016). Furthermore, most counties in western NY, one of the hardest hit regions of the state, were declared to be under natural disaster by the United States Department of Agriculture Farm Service Agency (USDA FSA) (Sweet et al. 2017). Even when drought is not extensive, NY is prone to short-term "flash" droughts which can greatly impact agricultural production (Otkin et al. 2018, Mo & Lettenmaier 2016).

There is a need to quantify the severity and the spatiotemporal extent of drought events, looking at a variety of drought related information across a spectrum of spatial and temporal scales. In 2006, Congress initiated the National Integrated Drought Information System (NIDIS) with a mandate to establish a national Drought Early Warning System (DEWS) through a network of regional DEWS (NE DEWS 2018). The Northeast Drought Early Warning System (NE DEWS) strategic plan established 4 priorities to enhance the quantification and awareness of drought in the Northeast. They are to: (1) enhance drought monitoring and research, (2) integrate and develop collaborative networks, (3) strengthen decision making to improve drought planning and preparedness, and (4) increase communication and utilization of drought and climate science (NE DEWS 2018). Priority 1 includes enhancing groundwater monitoring networks, expanding soil moisture monitoring in the region, strengthening citizen science networks, and developing forecast tools to predict future hydrologic conditions during periods of drought or potential drought. While efforts to address these priorities should eventually generate a more robust dataset with which to characterize drought extent and intensity, currently the U.S. Drought Monitor (USDM) is the most commonly employed tool for drought characterization in the U.S. The USDM, which has characterized the extent and intensity of drought in the U.S. since January 2000 and is jointly distributed through the National Drought Mitigation Center (NDMC) at the University of Nebraska-Lincoln, the National Oceanic and Atmospheric and Administration (NOAA), and the USDA (USDM 2019). The USDM uses a host of measured and historic data (e.g., precipitation, streamflow), drought indices (e.g., Palmer Drought Severity Index), satellite-based products (e.g., Vegetation Health Index (VHI)), and expert judgment to produce weekly maps on the severity of drought in the U.S. (USDM 2019, Palmer 1965, Svoboda et al.

2002). This data set encompasses precipitation, soil moisture, groundwater levels, streamflow, water supplies and a multitude of other variables related to drought (USDM 2019).

Two key issues with the USDM are the inconsistency in the input data and the inability to forecast future droughts. A variety of spatiotemporal scales are attributed to the input variables of the USDM, which are acquired from dissimilar sources (e.g., satellite based products, observational data, model output). There is a lack of consistency regarding which input variables are used to develop the USDM in specific regions, and the relative weights of these input variables used to produce the final USDM maps. For example, the VHI is a satellite based USDM input variable used to assess the impacts of drought on agricultural production. VHI outputs at two spatial resolutions, at 1 km and 16 km (NOAA 2019b). The USDM also employs climatological inputs (e.g., precipitation data from a local rain gauges or regional radar) (USDM 2019). The spatial resolution of the satellite-based products, precipitation data, and USGS streamflow data, all used to develop the weekly drought map, may not be consistent. This provides challenges to accurately define the borders of drought categories. In addition, the input variables used to develop the USDM sometimes vary across states. For example, in the Western U.S., additional climatological indicators, such as the Surface Water Supply Index (SWSI, for states where snowpack significantly impacts the water balance) or the Keech-Byram Drought Index (KBDI, for states where forest fires are prevalent), are incorporated into the development of the weekly drought maps (USDM 2019, Garen 1993, Varol & Ertuğrul 2016). In addition, the USDM only provides an assessment of the current drought status in the U.S.; it makes no determination of the potential for drought in the future.

To attempt to address both of these issues, the National Water Model (NWM) is proposed as a candidate tool to augment the informational sources that currently comprise the USDM. The NWM is developed collaboratively through a consortium of Universities (i.e. the University Corporation for Atmospheric Research (UCAR)), the NOAA National Weather Service (NWS) and the USGS (Sullivan 2016). The NWM runs at a relatively fine horizontal resolution (1 km) relative to some of the current USDM input

variables and provides forecasts at several different time intervals. Additionally, high resolution radar-based precipitation data can be used as an input to the NWM (NOAA 2019b). For these reasons, supplementing the USDM with NWM output may not only help improve the characterization of drought at finer horizontal resolutions, but should this model exhibit skill in characterizing drought, it might also aid in improved drought forecasting and the management of water resources during drought periods.

New York is home to a robust network of meteorological stations, the New York State Mesonet (Mesonet). Operated by the University at Albany, the Mesonet is a network of 126 spatially distributed meteorological stations that comprise the most advanced early warning weather protection system in the country (NY Mesonet 2019). The automated sensors at each station collect data on temperature, humidity, solar radiation, snow depth, soil moisture and temperature, and several other key climatological variables (NY Mesonet 2019). Both the NWM output and Mesonet soil moisture observations exist for the time period leading up to and continuing through the duration of the 2016 Northeast drought. Given the availability of data from the Mesonet and the NWM, New York provides a good location to explore the use of this data to improve drought characterization and potentially enhance the USDM.

After a brief description of the National Water Model, the following analysis focuses first on comparing NWM streamflow output to recorded streamflow from a set of unregulated USGS gauges in NY. A comparison is made of the streamflow magnitudes and the predicted drought category based on percentile ranges. Next, NWM soil moisture output is compared to soil moisture measurements at Mesonet sites, in addition to three longer term soil moisture stations across NY. Due to the short period of record at the Mesonet sites, and a somewhat incomplete record at the three longer term soil moisture stations, only magnitudes of observed soil moisture are compared; observational soil moisture datasets are not used to derive drought categories. Any data used to derive drought categories based on percentile ranges are additionally compared to those from the USDM, with a particular focus on the Northeast drought of 2016.

2.2 Description of National Water Model

The NWM is a tool to simulate current and forecast future water resources over the continental United States (CONUS) (Read et al. 2018). NWM analyses are categorized as: (1) Standard Analysis (a self-cycling 3-hour loop-back), (2) Long-Range Analysis (a 12-hour loop-back), and (3) Extended Analysis (a 28-hour loop-back, used to initialize the Standard Analysis) (Gochis et al. n.d., NOAA 2019c). From these analyses, the NWM is able to provide three forecasts: (1) Short-Range (18-hour deterministic (single value) forecast initialized by the Standard Analysis), (2) Mid-Range (a seven-member ensemble forecast initialized by the Standard Analysis), and (3) Long-Range (four-member ensemble forecast, initialized by the Long-Range Analysis) (Gochis et al. 2019, NOAA 2019c). The Short-Range forecast is executed hourly, while the Mid- and Long-Range forecasts are executed four times a day. For each forecast, streamflow is provided at roughly 2.7 million river reaches, and other hydrologic information (i.e. soil moisture) on either a 250 m or 1 km raster grid (Read et al. 2018, NOAA 2019c).

The NWM requires high-resolution radar-gauge precipitation data (e.g., Multi-Radar/Muti-Sensor System (MRMS) or Stage IV Multisensor Precipitation Estimator (MPE)), and forcing data from atmospheric modeling/assimilation systems (NOAA Rapid Refresh (RAP) or the High Resolution Rapid Refresh (HRRR)) or general circulation models (NOAA Global Forecasting System (GFS) (atmospheric coupled) or National Center for Environmental Protection Climate Forecasting System (NCEP CFS) (ocean atmospheric coupled)) (NOAA 2019a, 2019d, 2019e, 2019f, 2019h). United States Geological Survey (USGS) streamflow observations are also used as an input to assist in simulating current conditions and forecasts (NOAA 2019c).

The Weather Research and Forecasting Hydrological modeling system (WRF-Hydro) forms the core of the NWM's system (NOAA 2019c). WRF-Hydro is a distributed hydrologic model which determines vertical energy and moisture fluxes, and then the thermal and moisture states of soils (Powers et al. 2017,

Gochis et al. 2018). Modeling of sensible and latent heat, net radiation, interception, infiltration, infiltration-excess and percolation occurs first (Gochis et al. 2018). This information is then passed to either the sub-surface, or overland flow modules (Gochis et al. 2018). The sub-surface module typically consists of four zones from which lateral flow and the depth to the water table are determined (Gochis et al. 2018). After the sub-surface processes are modeled, WRF-Hydro then estimates overland flow. Overland flow occurs when the depth of water in any one grid cell exceeds a specific retention depth and is modeled by the diffusive wave equation with resistances established by Manning's roughness (Ogden 1997, Vieux 2001). Additionally, water can pass into lakes or reservoirs through a one-way process (i.e. water cannot move from lakes or reservoirs back into the landscape), while baseflow is determined through a conceptual catchment-storage "bucket" model (Gochis et al. 2018). For use with the NWM, WRF-Hydro is configured with Noah-MP Land Surface Model (LSM) to simulate the above processes.

2.3 Data and Methods

2.3.1 Data

The data employed in this analysis includes output from a 25-year NWM reanalysis (National Water Model v1.2 from http://edc.occ-data.org/nwm/getdata/), a variety of streamflow and soil moisture observational data, and USDM drought categories. Observational data includes USGS daily average streamflow, and Mesonet, NOAA, and USDA soil moisture observations. These data sets are described in the following sections.

2.3.1.1 NWM Output

All NWM output was obtained from a 25-year reanalysis period from 1993 to 2017. A reanalysis is one way to introduce dynamic consistency to model input variables, such that their spatial and temporal distributions are identical (Bengtsson et al. 2004). A reanalysis is useful to evaluate a modeling system's climatic bias/dynamic operating range in order to place future model anomalies into historical context. The National Weather Service (NWS) conducted a retroactive simulation of Version 1.2 of the NWM from January 1, 1993 to December 31, 2017 (NWM 2019); this reanalysis dataset is used in this study.

For each day of the simulation, both streamflow and soil moisture were obtained from the NWM reanalysis output at the 12:00 Coordinated Universal Time (UTC) timestep. NWM streamflow data was extracted from the nearest stream segment corresponding to the location of the USGS gauges used in this analysis. The stream segments were obtained from the NOAA Office of Water Protection map viewer (https://water.noaa.gov/map). NWM soil moisture output was extracted from the nearest grid cell corresponding to the location of a Mesonet site, from a 1 km raster grid across NY (NOAA 2019b). NWM soil moisture output has four layers at depths with thicknesses of 0-10 cm (SM1), 10-40 cm (SM2), 40-100 cm (SM3), and 100-200 cm (SM4). The top three (SM1, SM2 and SM3) correspond to the observational data used in this analysis (discussed below in Sections 2.3.1.2, and 2.3.1.3).

2.3.1.2 Streamflow Data

Xia et al. (2012), who compared NLDAS-2 simulated streamflow to observed streamflow at USGS gauging sites, employed 961 basins across the CONUS. These sites were chosen because they were generally small unregulated basins (drainage areas from 20 to 10,000 km^2) (Xia et al. 2012). Of the 961 sites used in Xia et al., 30 are located in NY and are used in this analysis. Characteristics for these sites may be found in Table 1, including the latitude and longitude of the sites, the feature ID of the stream segment from which NWM streamflow output was obtained, drainage area, and whether the period of record for the USGS gauging station encompasses the entire 25-year reanalysis period available for the NWM (January 1993 to December 2017). The drainage areas for the study sites range from 65.3 km^2 (USGS gauge #1414500 Mill Brook near Dunraven, NY) to 6491 km^2 (USGS gauge #1531000 Chemung River at Chemung, NY), while the majority of the study sites have drainage areas between 150 and 1,500 km^2.

Four sites (USGS #1530500 Newtown Creek at Elmira, NY; USGS #4214500 Buffalo Creek at Gardenville, NY; USGS #4217000 Towanda Creek at Batavia, NY; and USGS #4233000 Cayuga Inlet near Ithaca, NY) had incomplete records with respect to the 25-year NWM reanalysis period. The gauges for Newtown Creek and Cayuga Inlet were taken off-line during the reanalysis period and were moved to

new locations. Attempts were made to extend the record at these sites with concurrent and/or extended data taken from the new gauging location; however, the quality of the extended record at both sites appeared to be poor. The gauge at Newtown Creek was taken off-line in 2008 and the gauge at the Cayuga Inlet was taken off-line in 2012. Because this analysis places particular emphasis on the Northeast drought of 2016, and information from these sites is not concurrent with that period, these sites were removed from the analysis.

From September 30th, 1998 through September 29th, 1999 the gauging station at Towanda Creek was not actively recording data, and from September 30th, 2007 through September 30th, 2012 the gauging station at Buffalo Creek was not actively recording data. At both of these sites, no attempts were made to fill the record gaps; instead, at these sites the NWM simulations were subsetted to be concurrent with the USGS streamflow observations.

2.3.1.3 Soil Moisture Data

The 126 NY Mesonet stations became active from late 2015 to mid 2017, at which point the network became fully operational. No data appears to be available from the Mesonet station at Stony Brook (located at Three Village Central School District). Additionally, no NWM soil moisture output was available for the Raquette Lake Mesonet station located at Camp Huntington; the NWM location is identified as a water body (Raquette Lake). The Mesonet stations in the boroughs of New York City (Brooklyn located at Brooklyn College, Bronx located at Lehman College, Manhattan located at Hunter College CUNY, Queens located at Queens College, and Staten Island located at CUNY Staten Island) do not appear to collect soil moisture data. These are rooftop sites in urban areas with no soil moisture measures. The remaining 119 Mesonet stations are used in this analysis. Soil moisture at the Mesonet sites is generally observed at depths of 5 cm, 25 cm and 50 cm (unless bedrock limits the installation of deeper probes), which correspond to the midpoints of the SM1, SM2 and SM3 NWM soil moisture layers. At five of the 119 sites in consideration (Copenhagen located at Williams Farm, Deposit located at NYC

DEP Deposit, Fayetteville located at Green Lakes State Park, Grove located at Johnsen Property, and Roxbury located at NYC DEP Fanny Brook Road), no soil moisture was recorded at the 50 cm depth (due to bedrock limitations or a shallow water table), which corresponds to the midpoint of the SM3 NWM soil moisture layer. For these five sites, only soil moisture readings at depths of 5 cm and 25 cm (which correspond to the SM1 and SM2 NWM soil moisture layers) are considered. Mesonet soil moisture data is recorded on a 5-minute interval, and daily instantaneous data was obtained using the reading at 12:00 UTC.

One problem with the Mesonet soil moisture data is its short period of record that is not concurrent with the entire 25-year NWM reanalysis period. A smaller number of longer-term soil moisture stations are also employed in this analysis. The NOAA Climate Reference Network (CRN) has two sites in NY: one at Millbrook, NY (eastern NY), and one in Ithaca, NY (west-central NY) (NOAA 2019g). The CRN takes soil moisture readings at depths of 5", 10", 20", 50", and 100". Depths of 10" (25 cm) and 20" (~50 cm) roughly correspond to the midpoints of the SM2 and SM3 NWM soil moisture layers and are similar to the deeper two Mesonet soil moisture depths. The two CRN sites became operational in late 2004, recording soil moisture at sub-hourly, hourly, and daily timesteps. For this analysis, the daily data was obtained as the hourly reading at the 12:00 UTC timestep (which encompasses data recorded from 11:05 UTC to 12:00 UTC). The USDA Soil Climate Analysis Network (SCAN) has one site at Geneva, NY (western NY) (USDA 2019). The SCAN takes soil moisture readings at depths of 2", 4", 8", 20", and 40". Depths of 2" (5 cm), 8" (20 cm), and 20" (~50 cm) roughly correspond to the midpoints of the SM1, SM2, and SM3 NWM soil moisture layers and are similar to each of the Mesonet soil moisture depths. The Geneva SCAN site became active for one month in October 1994, and then resumed uninterrupted data collection from October 1995 to the present. As such, the "trial" month in 1994 is not considered for this analysis. The SCAN network records data at the hourly and daily timesteps. SCAN records 4 hourly readings over a day at 12:00am, 6:00am, 12:00pm, and 6:00pm Local Standard Time (LST).

Table 1- Watershed characteristics for 30 USGS gauged streamflow sites, including the USGS gauge number, the Feature ID of the National Water Model stream, latitude, longitude, drainage area and whether the historic record is complete from 1993 to 2017.

USGS Gauge Number	NWM Feature ID	Latitude	Longitude	Drainage Area (km^2)	Complete Record
01321000	22297114	43.3527778	-74.270278	1271.69	Yes
01334500	22290307	42.9386111	-73.376944	1320.90	Yes
01350000	3247466	42.3194444	-74.436667	613.83	Yes
01365000	6200212	41.8663889	-74.487222	99.20	Yes
01371500	6200432	41.6861111	-74.165278	1800.04	Yes
01372500	6212050	41.6530556	-73.8725	468.79	Yes
01413500	1748583	42.1447222	-74.653611	422.17	Yes
01414500	1748611	42.1061111	-74.730556	65.27	Yes
01415000	1748589	42.12	-74.818611	85.99	Yes
01420500	1750585	41.9464167	-74.979667	624.19	Yes
01421000	1752137	41.9735556	-75.175194	2030.55	Yes
01423000	2613578	42.1661111	-75.140028	859.88	Yes
01426500	2614210	42.0030278	-75.383528	1541.04	Yes
01435000	4147956	41.89	-74.59	172.49	Yes
01503000	8088149	42.0352778	-75.803056	5780.86	Yes
01512500	9423965	42.2180556	-75.848333	3840.95	Yes
01521500	8110879	42.3958333	-77.711389	79.25	Yes
01530500	8117963	42.1044444	-76.798056	200.72	No
01531000	8118503	42.0022222	-76.634722	6490.51	Yes
03011020	8969860	42.1563889	-78.715278	4164.70	Yes
04214500	15576299	42.8547222	-78.755	367.78	No
04215500	15576309	42.8297222	-78.775	349.65	Yes
04217000	15568739	42.9975	-78.188611	442.89	No
04221000	15550135	42.1222222	-77.957222	745.92	Yes
04223000	15549241	42.5702778	-78.042222	2548.55	Yes
04230500	15537859	43.01	-77.791389	518.0	Yes
04233000	21983265	42.3930556	-76.545	91.17	No
04234000	21983115	42.4533333	-76.472778	326.34	Yes
04243500	22024054	43.0975556	-75.639417	292.67	Yes
04262500	15491848	44.1855556	-75.330833	668.22	Yes

LST is 5 hours behind UTC; therefore, the SCAN observation at 6:00am LST (the closest to 12:00 UTC) were taken as the daily reading. These longer-term soil moisture datasets were also determined to be problematic. The sensors at these stations were commonly inactive which resulted in an overall incomplete record. While both CRN sites contained a complete record (at the daily timestep), the SCAN site was missing 191 days entirely. The SCAN site was the most complete, while the CRN sites were inactive for roughly half of their total observational record. See Table 2 below for a detailed summary of the missing and inactive days for the CRN and SCAN sites.

Table 2- Long-Term Soil Moisture Missing Days and Non-Recording Days

			Non-Recording Days (Missing or Otherwise)		
Site	Network	Missing Days	SM1	SM2	SM3
Geneva	SCAN	191	535 (~1.5 years)	618 (~2 years)	2698 (~8-years)
Ithaca	CRN	0	n/a	2755 (~7.5 years)	2648 (~7 years)
Millbrook	CRN	0	n/a	2816 (~8 years)	2799 (~7.5 years)

2.3.2 Methods

A numerical comparison of observational and NWM simulated data is conducted, followed by an assessment of the skill in this data at reproducing USDM drought categories at the county scale. A host of similar, yet slightly different statistics are calculated for the streamflow and soil moisture by comparing NWM simulation output to observational data. Both USGS observations and NWM simulated streamflow are then used to derive drought categories and the bias and root mean square error between the drought categories derived from NWM data and the USDM observed drought categories were determined for 19 counties in NY. Similarly, NWM soil moisture in the SM1, SM2, and SM3 layers is used to derive the drought categories, and the bias and RMSE were determined for all 62 of NY's counties. Finally, a drought map is produced by this data and is compared to the USDM map for September 2016, the height of the 2016 drought in NY. These analyses are explained in the following sections of this thesis.

2.3.2.1 Comparison Methodology

Table 3 contains a list of the comparisons made between NWM output and observational data. For each concurrent record, a number of numerical comparison metrics were calculated to assess the ability of the NWM output to represent streamflow and soil moisture observations. For NWM output and USGS streamflow observations, an assessment was also made to determine their skill in identifying locations that are within a specific USDM drought category, and how these drought categories compare with drought categories based on historic USDM maps.

2.3.2.2 Numerical Comparison

At each of the 28 USGS NY gauging stations with historic streamflow records, 7-day, 14-day, and 28-day averages, a common drought averaging period employed by the USGS (USGS 2018), were compared to NWM output at the associated NWM stream segment. Averages were non-overlapping (i.e. 7-day

Table 3- Comparisons made between NWM output and recorded streamflow and soil moisture values.

NWM Output	Comparison	Location	Number of Sites	Record Length
Streamflow	USGS historic observations	USGS Gauge Location	28	20 – 25 years (1993-2017)
Soil Moisture	Mesonet	Mesonet locations: 5, 25 and 50 cm	114	2 – 3 years (2015-2018)
Soil Moisture	Mesonet	Mesonet locations: 5, 25 cm	5	2 – 3 years (2015-2018)
Soil Moisture	USDA CRN	CRN locations: 25 and 50 cm	2	14 years (2004-2018)
Soil Moisture	NOAA SCAN	SCAN locations: 2, 25 and 50 cm.	1	23 years (1995-2018)

averages are for days 1 – 7, 8 – 14, etc.) and calculated with the USGS daily average streamflow observations and the NWM daily data using the day of an observation and the preceding days. For each averaging period, comparison metrics were determined. The relative bias was calculated as:

$$RBias = \frac{\sum_{i=1}^{N} \frac{(\hat{\theta}_i - \theta_i)}{\theta_i}}{N} \quad (1)$$

where $\hat{\theta}_i$ is the averaged NWM streamflow output on day i, θ_i is the averaged USGS streamflow observation on day i, and N is the number of timesteps (days) compared. RBias can be negative or positive, and estimators are generally preferred if RBias is closer to zero. While bias is a common metric of systematic error, an estimator can have a low bias yet be imprecise if over- and under-estimation errors cancel each other. The average relative absolute deviation (ARAD) was also calculated as:

$$ARAD = \frac{\sum_{i=1}^{N} |\frac{(\hat{\theta}_i - \theta_i)}{\theta_i}|}{N} \quad (2)$$

ARAD is strictly positive, with values closer to 0 representing better model fits. 100*ARAD represents the average percent error, which is relatively easy to interpret. To assess the precision of the estimators, the relative root mean square error (RRMSE) was calculated as:

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right)^2}{N}} \quad (3)$$

Use of relative metrics (RBias, ARAD, and RRMSE) mitigates some of the impact of different streamflow magnitudes on these metrics, allowing for fairer comparison across sites.

While it presents information similar to the RRMSE, the Nash-Sutcliffe Efficiency, a widely used performance metric, was also calculated in real- and log-space:

$$NSE = 1 - \left(\frac{\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^{N} (\theta_i - \bar{\theta})^2}\right) \quad (4)$$

$$LNSE = 1 - \left(\frac{\sum_{i=1}^{N} (\log(\hat{\theta}_i) - \log(\theta_i))^2}{\sum_{i=1}^{N} (\log(\theta_i) - \bar{\theta}_L)^2}\right) \quad (5)$$

where $\bar{\theta}$ is the mean of the θ_i and $\bar{\theta}_L$ is the mean of $\log(\theta_i)$. The real-space NSE is highly sensitive to larger observations, especially in series with high coefficients of variation, and can be upwardly biased similar to the correlation coefficient (Stedinger 1981), while the log-space NSE places more emphasis on the lowest observations. For streamflow, the above metrics were calculated for the entire record as well as

for data below the 30th percentile of θ_i (USGS x-day streamflow averages), allowing for assessment NWM performance during the driest periods of the record.

Similar comparison metrics were also calculated for soil moisture series, though relative metrics were not employed because soil moisture is constrained to be in the range from 0 to 1, although given that the soil moisture data was obtained as a volumetric water content (VWC) the expected range for the soil moisture is 0.05-0.45. 7-day, 14-day, and 28-day non-overlapping averages were determined for the observations at the Mesonet sites, the 3 long-term soil moisture records and the NWM soil moisture output. The three averaging periods were calculated for applicable data in the SM1, SM2, and SM3 soil moisture layers. For each averaging period, at each soil moisture layer the bias was calculated as:

$$Bias = \frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)}{N} \quad (6)$$

where $\hat{\theta}_i$ is the 7-, 14-, or 28-day aggregated NWM soil moisture output on day i, θ_i is the measured aggregated soil moisture on day i, and N is the number of timesteps (days) compared. The root mean square error of the NWM averaged soil moisture was calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}{N}} \quad (7)$$

The average absolute deviation was also calculated as:

$$AAD = \frac{\sum_{i=1}^{N}|\hat{\theta}_i - \theta_i|}{N} \quad (8)$$

NSE and LNSE were also calculated for soil moisture. For soil moisture, the above metrics were calculated for all days across all concurrent years.

2.3.2.3 Drought Classification

The intensity of drought in New York was then classified according to NWM streamflow and soil moisture output (at the SM1, SM2, and SM3 layers) and USGS streamflow observations (28 sites) using USDM drought percentiles. For each of these datasets, 7-day non-overlapping averages were calculated

for all days in all years starting with the beginning of the reanalysis period. The USDM produces drought maps on a weekly basis, so in this initial analysis 7-day averages were employed. 14- and 28-day averages were also computed, and compared against the corresponding day in which the USDM released its map. For the longer averaging periods, data extending prior the period the USDM considers is used. The goal was to determine the effect a given averaging period had on the replication of USDM drought categories. For each 7-day average on a specific day of a specific year, the non-exceedance probability was calculated considering each 7-day window to be an independent sample (i.e. January 7th is considered separately from January 14th etc.) with a sample size, *n*, that is the number of years of observations over the reanalysis period (generally n= 25, or n=24 from January 1-6). The non-exceedance probability of a 7-day average was then determined using a Weibull plotting position:

$$pp = \frac{i}{n+1} \quad (9)$$

where i is the rank of the 7-day average (i = 1 for the smallest observation and n for the largest) (Weibull 1939). The Weibull plotting position produces unbiased estimators of non-exceedance probabilities (Stedinger et al. 1993). This was performed similarly for the 14- and 28-day averaging periods.

Non-exceedance probabilities were calculated using the concurrent records for NWM streamflow output and USGS records at all 28 NY gauging stations. Using a similar procedure, non-exceedance probabilities were also calculated for 7-day average NWM soil moisture output for SM1, SM2 and SM3 layers. Soil moisture observations at the Mesonet and at each of the three long-term stations CRN and SCAN were not used to derive drought categories. As previously stated, the Mesonet suffered from a short period of record, while the longer-term soil moisture records contained many gaps. These conditions lead to complications in the analysis and inconsistencies in determining non-exceedance probabilities.

The Mesonet has at least one station in each of the 62 counties in NY. For counties with more than one site, an average of the NWM soil moisture non-exceedance percentile from each Mesonet station was taken, and then classified using USDM drought categories. The NWM provides a spatially complete dataset of soil moisture across CONUS (i.e. there are no gaps). This experiment was simplified by only considering NWM output in grid cells corresponding to the locations of Mesonet stations. NWM and USGS streamflow were compared to a smaller subset of counties in NY as the 28 sites did not encompass all counties in the state. Similarly, if multiple sites existed in any one county, an average of the non-exceedance percentile was taken. Ultimately, 19 counties were compared for the streamflow data available. The at-site data was used to conduct a qualitative assessment of drought classification in NY, while the county averaged data was used to conduct a quantitative assessment of drought classification in NY.

The USDM classifies drought as No Drought (ND) ($> 30^{th}$ percentile), D0 ($\leq 30^{th}$ percentile), D1 ($\leq 20^{th}$ percentile), D2 ($\leq 10^{th}$ percentile), D3 ($\leq 5^{th}$ percentile), and D4 ($\leq 2^{nd}$ percentile). Based on the non-exceedance probabilities calculated for the streamflow and soil moisture series, drought classifications according to each type of data were obtained, and drought maps for the state of New York were developed. For the streamflow-based maps, the boundaries of the 28 watersheds were plotted and filled accordingly to their classification. For the soil moisture-based maps, points corresponding to the location of the 119 Mesonet stations were plotted across New York State. These points were filled in a similar manner based on the classification of drought from the NWM soil moisture output at the location of the nearest grid cell to the respective Mesonet station. These maps were developed in R, and a color scheme which mimics that of maps published by the USDM was developed to provide a qualitative assessment of how well our classification scheme matches the USDM at each of the streamflow or soil moisture sites.

A quantitative assessment of our drought classification was also conducted. Data from the USDM was obtained at the county scale from the USDM (https://droughtmonitor.unl.edu/Data/DataDownload/

ComprehensiveStatistics.aspx). The USDM reports the percentage of the county at or greater than a specific drought classification. From this, percentages of the county classified within a specific non-exceedance probability range was determined. ND was given a numeric value of -1, only D0 (30% ≤ pp < 20%) a value of 0, only D1 (20% ≤ pp < 10%) a value of 1, only D2 (10% ≤ pp < 5%) a value of 2, only D3 (5% ≤ pp < 2%) a value of 3, and only D4 (pp ≤ 2%) a value of 4. Both a weighted average of all categories a county was classified, and the category at which the majority of the county was classified, were determined and used to compare against NWM output and USGS streamflow observations. 7-day averages where both the USDM and any one of our data sources (NWM soil moisture or streamflow, and USGS streamflow observations) were classified as ND were removed (this would be days where both the USDM and the data source indicated there was no drought). For each dataset, this resulted in removing roughly half of the pairs. For each county available (19 for streamflow and 62 for soil moisture), the bias and RMSE were determined between the drought categories derived from the NWM streamflow output, NWM soil moisture output or USGS streamflow observations and the USDM drought categories. This was completed in a similar manner for the longer averaging periods.

2.4 Results and Discussion

The results and potential implications of our findings are presented in the following sections. The numerical streamflow statistics are presented and discussed first. This is proceeded by the results for the soil moisture statistics. Finally, we present the qualitative and quantitative assessment of how well NWM output produces drought classifications.

2.4.1 Streamflow

At each of the 28 sites, non-overlapping 7-, 14- and 28-day average streamflow from the NWM is compared to the USGS gauged record for the concurrent period of record. At each site, the RBIAS, ARAD, RRMSE, NSE and LNSE were calculated. Figures 1 - 5 present the RBIAS, ARAD, RRMSE, NSE and LNSE for 7-day averages. Each of these figures contains two boxplots: one for the entire period

of record (ER), and one for when the USGS 7-day average streamflow was below the 30th percentile (30P). Since the results for the 14- and 28-day averages were similar, they are presented in Appendix 2.

From Figure 1, across the streamflow sites in NY, the NWM streamflow output generally had a positive RBias. For the ER, 20 sites (71%) had an RBias greater than zero, and for 30P 27 sites (96%) had a positive RBias. The RBias for the ER ranged from -23% to 129% across all sites (mean 19%), while for the 30P this range increases shifted upward from -4% to 295% (mean 70%). This indicates that the NWM generally overestimates 7-day average streamflow, and this overestimation generally increases for the lowest 7-day average flows. Note that results for 14- and 28-day averages were similar to those for 7-day averages, though the NWM was in greater agreement when the averaging period increased (see Figures in Appendix 2). Figure 2 contains the ARAD, a measure of the relative error, for 7-day averages. The ARAD ranged from 30% to 149% (mean 50%) for ER, and from 30% to 296% (mean 80%) for 30P. This indicates that the NWM 7-day average streamflow at best has a 30% error and is generally over 60% for the lowest 7-day averages. Figures 3 and 4 present the NSE and LNSE for the 7-day average streamflows. NSE is similar to the coefficient of determination and represents the percentage of variability in the USGS 7-day averages explained by the NWM. The NSE is very sensitive to large observations; the LNSE places more emphasis on smaller observations. For the ER, the NSE of 7-day averages ranges from 0.19 to 0.77 and the LNSE ranged from -0.04 to 0.82; for 30P, the NSE ranges from -73 to -0.21 and the LNSE ranged from -6.1 to 0.29. Examining the 30P results, 50% of the sites had an NSE below -3.4 while 50% of the sites had an LNSE below -0.14. While the NSE for the ER appears to do well at producing 7-day averages at some sites, overall there appears to be relatively poor performance of this model, and performance decreases considerably when examining the lowest 7-day averages (30P). Figure 5 contains the RRMSE for the 7-day averages. RRMSE ranged from 0.398cfs to 2.39cfs for the ER and ranged from 0.446 to 2.89 for the 30P. The extremely high RRMSE at some sites indicates that occasionally the NWM 7-day average streamflow is much different than the recorded USGS value. Plots for NSE, LNSE, and RRBIAS for 14- and 28-day streamflow averages can be found in Appendix 2.

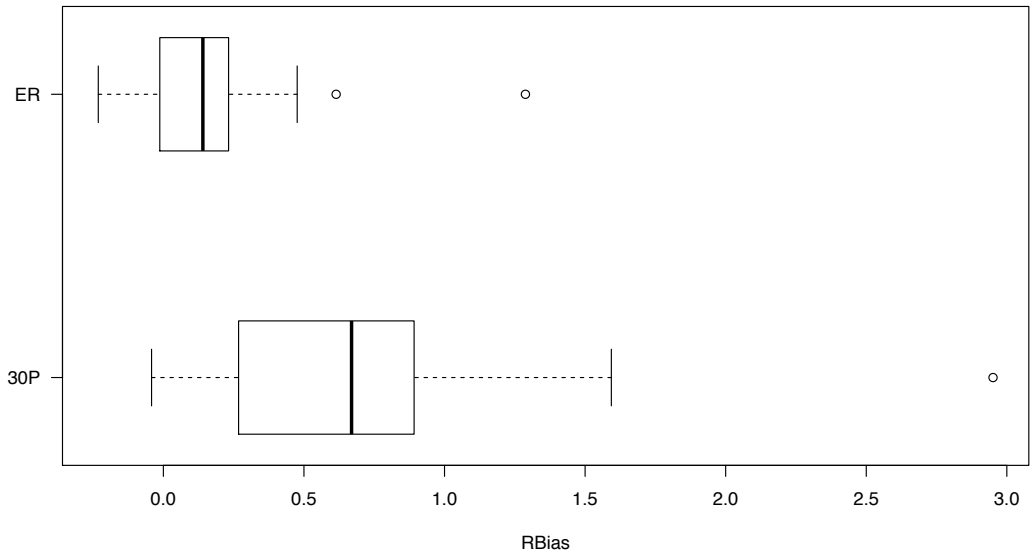


Figure 1- Relative Bias (RBias) for non-overlapping 7-day averages at the 28 USGS gauging stations for entire record (ER) and for less than the 30th percentile (30P).

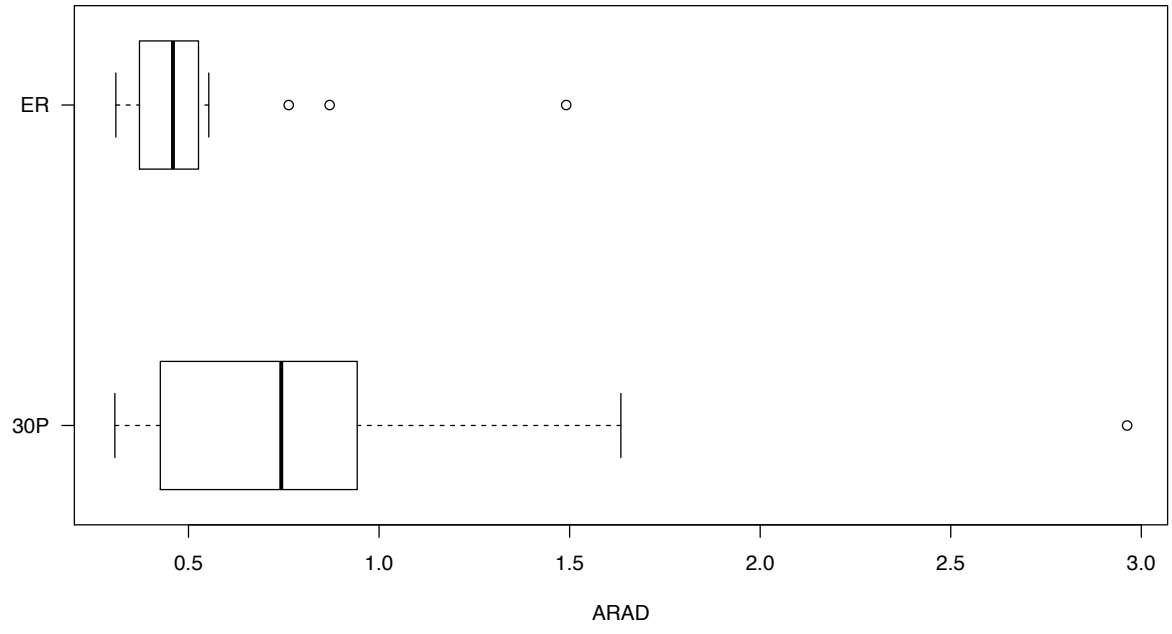


Figure 2- Average Relative Absolute Deviation (ARAD) for non-overlapping 7-day averages at the 28 USGS gauging stations for entire record (ER) and for less than the 30th percentile (30P).

In general, the results for 14- and 28-day averages were similar to 7-day averages. As with RBIAS and ARAD, when the averaging period increases, there is a slight increase in the performance of the streamflow estimators.

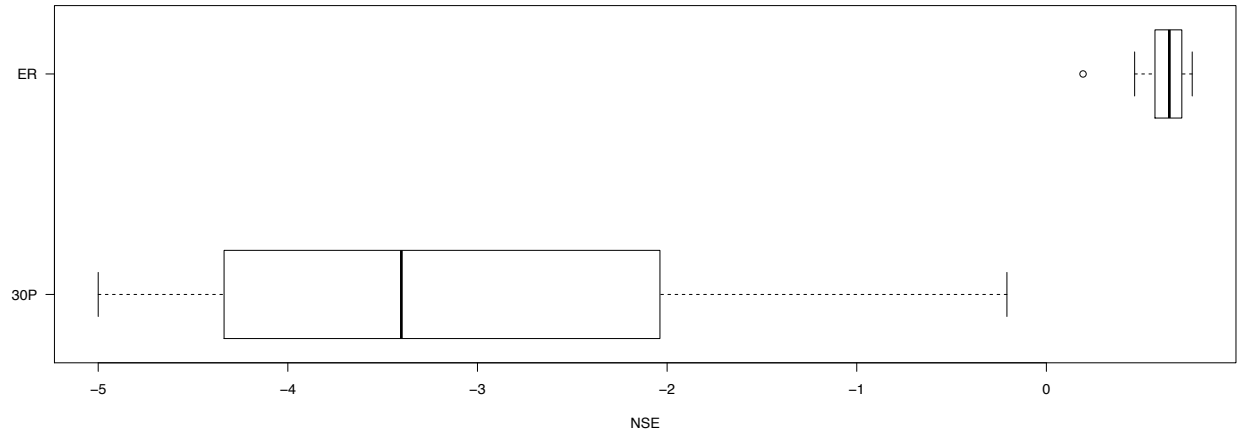


Figure 3- Nash Sutcliffe Efficiency (NSE) for non-overlapping 7-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

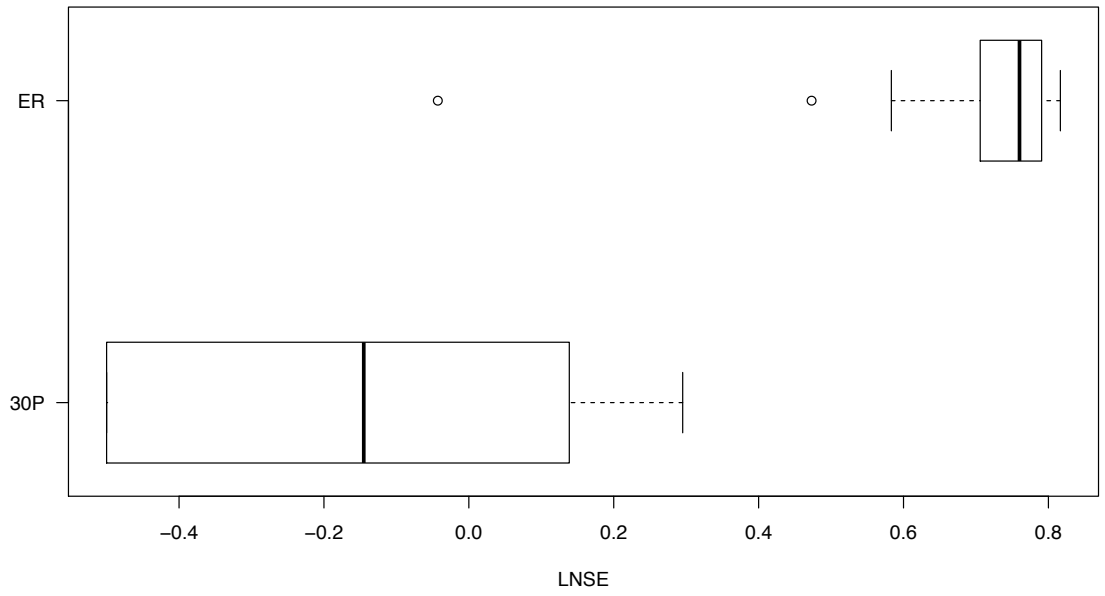


Figure 4- Log-space Nash Sutcliffe Efficiency (LNSE) for non-overlapping 7-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

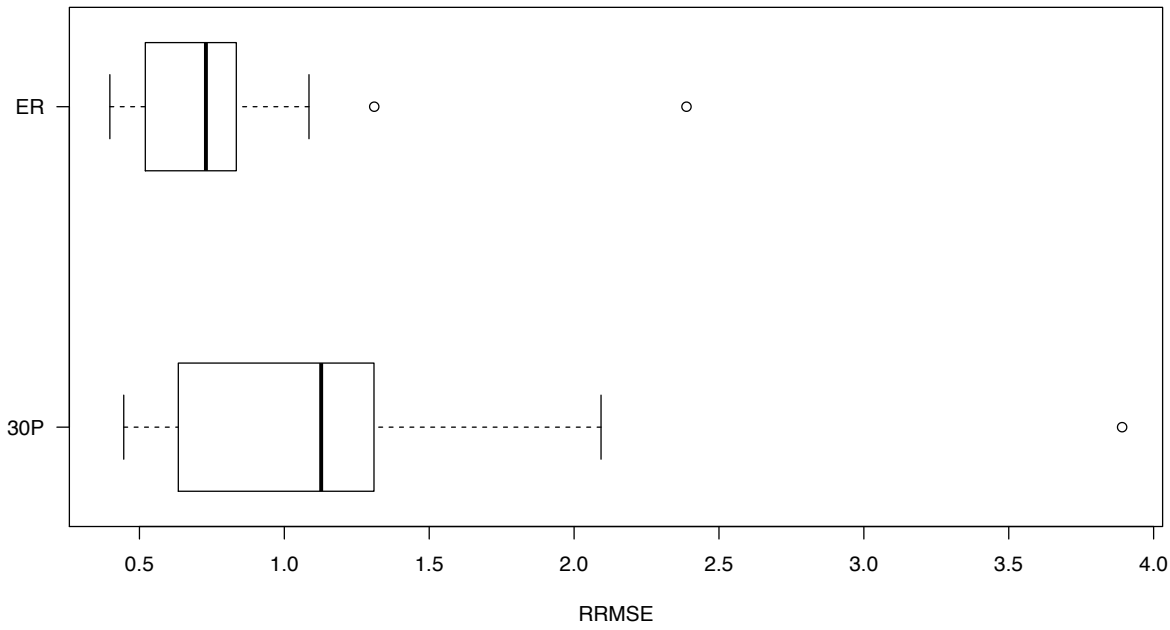


Figure 5- Relative Root Mean Square Error (RRMSE) for non-overlapping 7-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

To further assess the skill of the NWM streamflow, a set of 3 sites were chosen based on their ARAD values. The site with the lowest ARAD (USGS #01334500 Hoosic River near Eagle Bridge NY, ARAD = 31%), the site with the highest ARAD (USGS #01521500 Caniseto River at Arkport NY, ARAD = 150%), and the site closest to the median (USGS #04223000 Genesee River at Portageville NY, ARAD = 45%) were selected. For each site, the NWM 7-day average streamflow output was plotted versus the USGS 7-day average streamflow observations. The lowest 30% of those observations are highlighted blue. Additionally, each plot contains a regression line and a 45-degree line (which would indicate perfect agreement). Generally, as the ARAD increases, the departure of the regression line from the 45-degree line increases. At all sites, the regression line slope positively departs from the 45-degree line, indicating a slight over-prediction of the NWM. The site with the worst ARAD has the largest departure, though there is a large number of high outliers (see Figure 7). It appears the NWM generally overestimates 7-day average streamflow at the smaller streamflow values, and underestimates 7-day average streamflow at the larger streamflow values; such behavior (reduced variability) is intrinsic in models which do not add error

back to predictions (Farmer & Vogel 2016). The Hoosic River (watershed area 1320 km^2) is located in northern NY, while both the Caniseto River (watershed area 79 km^2) and Genessee River (watershed area 2548 km^2) are located in western NY, about 25 miles apart from each other and in the region of New York hit hardest by the 2016 drought.

2.4.2 Soil Moisture

At each of the 119 sites, non-overlapping 7-, 14- and 28- day averages were determined for soil moisture output from the NWM and is compared to Mesonet observed soil moisture record for the concurrent period of record in the SM1 and SM2 layers. In the SM3 layers, only 114 sites were considered as 5 sites were found to be non-recording at that depth. At each site, for in each profile, the Bias, AAD, RMSE, NSE and LNSE were calculated. Figures 7 – 11 present the Bias AAD, RMSE, NSE and LNSE for 7-day averages. Each of these figures contains three boxplots: one for each soil moisture profile in consideration, SM1, SM2, and SM3. A comparison of NWM soil moisture to the CRN and SCAN sites is also presented. The same host of statistics is determined (Bias, AAD, NSE, LNSE and RMSE) for each site at each available depth. These results are marked on their respective boxplots accordingly. Similarly, for both the Mesonet and long-term soil moisture comparisons, results for the 14- and 28-day averages were determined to be similar, and are presented in Appendix 2.

Figure 7 contains the Bias for the NWM soil moisture values at three depths (SM1, SM2, and SM3). Across all sites, the bias for all three soil moisture layers appears to indicate that on average the NWM is a very poor predictor of soil moisture. In the SM1 layer when looking at Mesonet data the bias has a range of -0.38 to 0.26 (mean -0.016), and 50% of the sites have a bias between -0.11 and 0.06 (the interquartile range). Considering that the soil moisture is reported as a VWC, this is an incredibly large range, one that is likely physically infeasible. Results for SM2 and SM3 layers are similar to those of SM1, and although there appears to be a slight improvement in performance for the deeper soil moisture. When comparing to the CRN and SCAN sites, the NWM tends to overpredict soil moisture. In the SM1

layer, the bias was 0.062 at the Geneva SCAN site. In the SM2 layer, the bias ranges from 0.06 (Ithaca CRN) to 0.1 (Geneva SCAN), and in the SM3 layer the bias ranges from 0.06 (Ithaca CRN) to 0.12 (Millbrook CRN). Figure 8 contains the AAD for the NWM soil moisture. While at a small number of sites the ARAD is close to zero, in general 50% of the sites have an AAD greater than 0.09. When comparing to the long-term records, in the SM1 layer the AAD was determined to be 0.065. In the SM2 layer, the AAD ranged from 0.06 Ithaca CRN) to 0.1 (Geneva SCAN), and in the SM3 layer the AAD ranged from 0.06 (Ithaca CRN) to 1.2 (Millbrook CRN). As soil moisture measurements are constrained over a relatively small range, these appear to be extremely large errors.

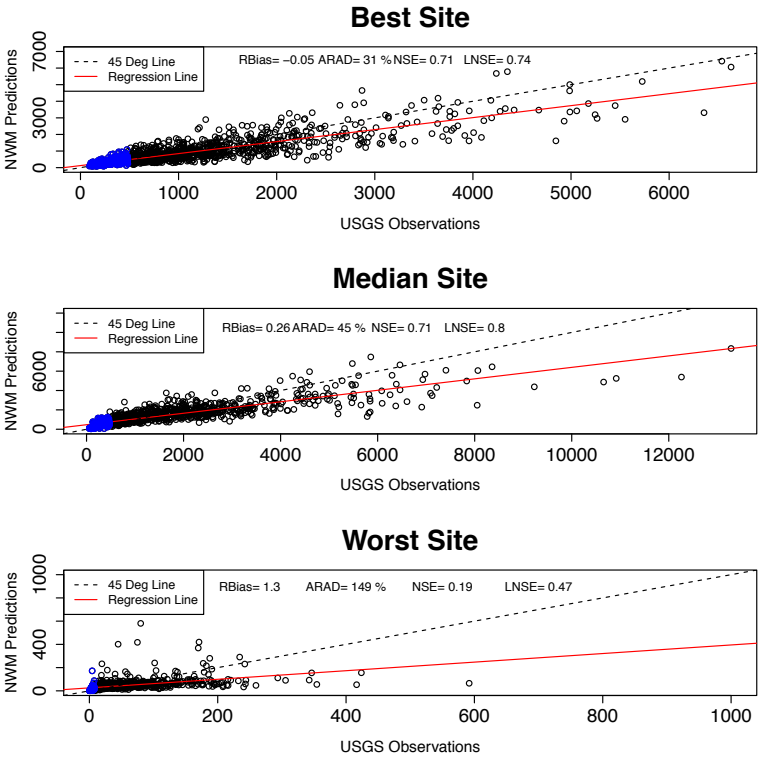


Figure 6- Plots of NWM 7-day average streamflow versus USGS observed 7-day average streamflow at site with the lowest ARAD, median ARAD, and highest ARAD.

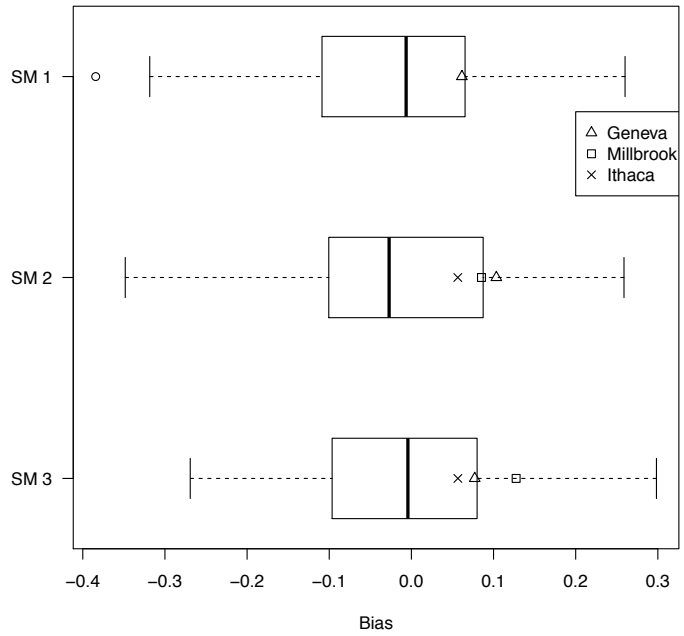


Figure 7- Bias in the SM1, SM2, and SM3 7-day average NWM soil moisture values compared to Mesonet (boxplots) and CRN and SCAN (symbols) sites.

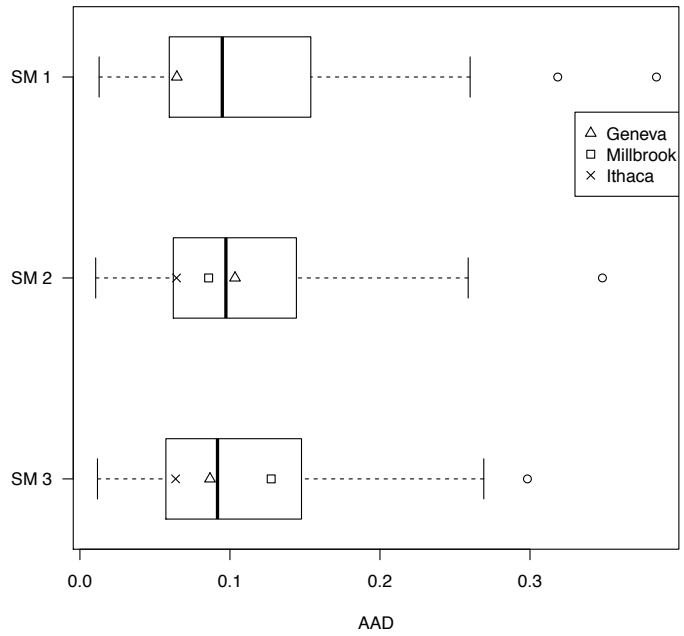


Figure 8- Average Absolute Different (AAD) in the SM1, SM2, and SM3 7-day average NWM soil moisture values compared to Mesonet (boxplots) and CRN and SCAN (symbols) sites.

Similarly, the NSE and LNSE in Figures 9 and 10 indicate very poor performance for the NWM soil moisture values compared to the Mesonet. In these figures, all sites with an NSE or LNSE < -1 were plotted as -1 on the figure. Table 4 contains the number of sites where this occurs for each averaging period considered (for the 7-day averages: ~50% of the sites for SM1 and SM2; more for SM3). Note that an NSE of 0 indicates the model is doing no better than the mean of the observations; over 75% of the sites have NSE less than 0. An NSE less than -1 indicates a major issue. There were similarly poor results when comparing against the longer-term soil moisture records. The LNSE was never greater than 0 and was greater than -1 in the SM2 layer at the Geneva SCAN site and in the SM3 layer at the Millbrook CRN site. The only positive NSE value was determined in the SM2 layer at the Ithaca CRN site. The NSE was less than -1 in the SM2 layer at the Geneva SCAN and Millbrook CRN sites, and in the SM3 layer at the Millbrook CRN site. 50% of the RMSE for SM1 were in the range 0.07 to 0.17; SM2 and SM3 had similar interquartile ranges (see Figure 11). Comparison to the longer-term soil moisture demonstrated similarly high RMSE values with ranges in the SM1 and SM3 layers from roughly 0.08 to 0.13. Again, the NWM seemed to perform slightly better with increasing depth. Similar to streamflow, the RMSEs appear large in relationship to the magnitude of the soil moisture, a further indication of the inability of the NWM soil moisture to reproduce the observed soil moisture at the Mesonet sites.

Table 4- Number of Sites in each soil moisture profile with NSE and LNSE values below -1.

Aggregation	Statistic	SM1	SM2	SM3
7-day	NSE	57	77	81
	LNSE	52	67	77
14-day	NSE	63	81	82
	LNSE	58	75	80
28-day	NSE	68	77	88
	LNSE	65	63	83

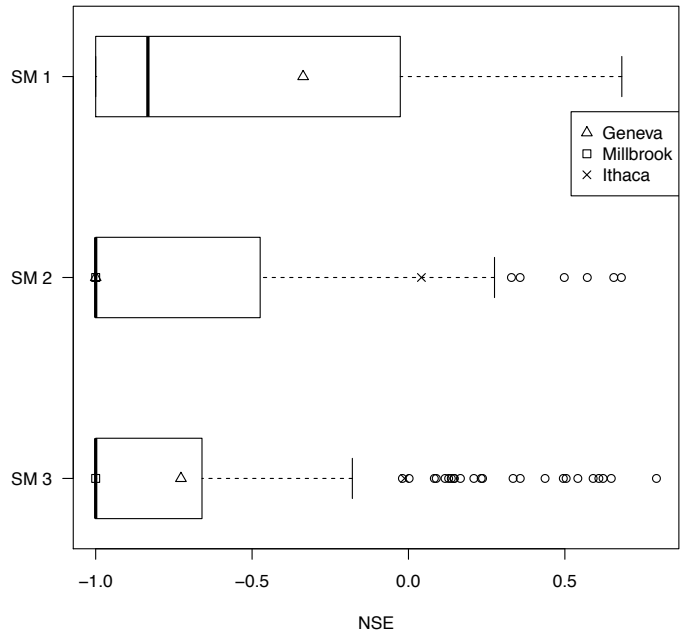


Figure 9- NSE in the SM1, SM2, and SM3 7-day average NWM soil moisture values compared to Mesonet (boxplots) and CRN and SCAN (symbols) sites.

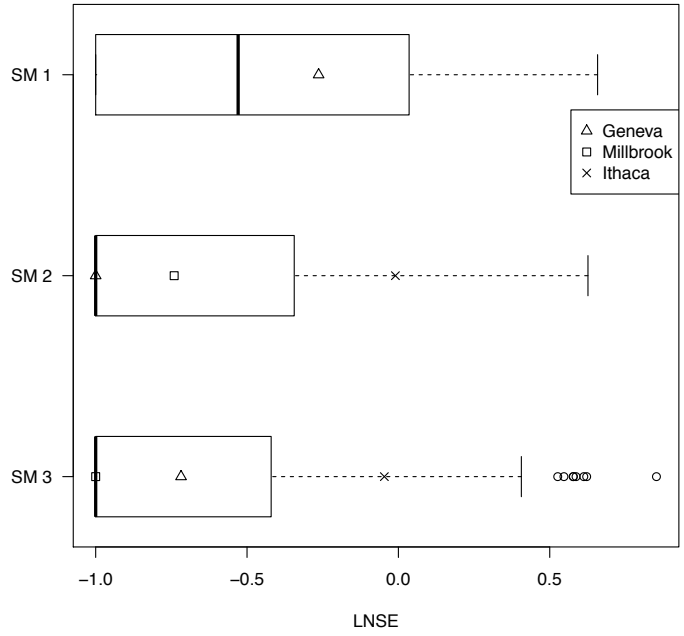


Figure 10- LNSE in the SM1, SM2, and SM3 7-day average NWM soil moisture values compared to Mesonet (boxplots) and CRN and SCAN (symbols) sites.

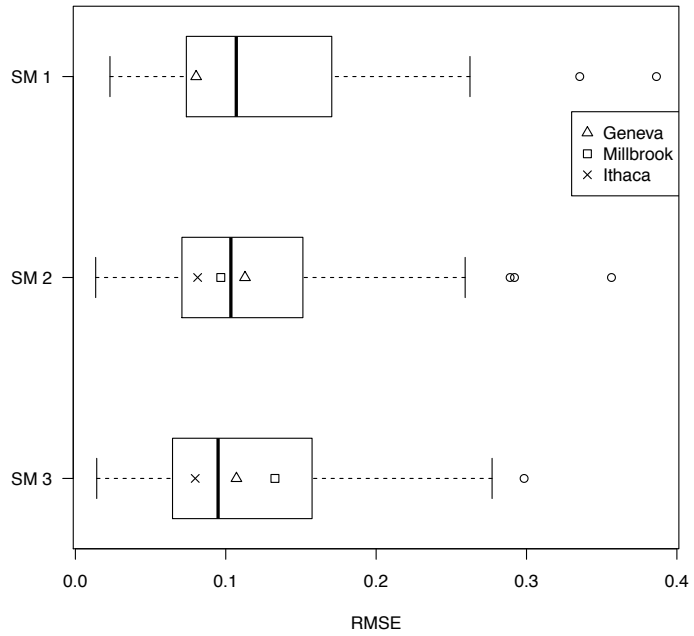


Figure 11- RMSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 7-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

Similar to the analysis performed for streamflow, the sites with the smallest, largest and AAD nearest to the median AAD in the SM1 layer were selected for further examination. Figure 12 contains NWM 7-day average in the SM1 layer soil moisture versus Mesonet 7-day average soil moisture at Newcomb, NY (ARAD = 1.28%), Andes, NY (ARAD = 9.49%), and Old Forge, NY (ARAD = 38.4%); the plots also contain a regression line and a 45-degree line. Across these three sites, the NWM is doing a very poor job reproducing the observed Mesonet soil moisture, missing badly at Andes and Old Forge. In addition, at Andes and Old Forge, while the Mesonet observations have a range of about 0.1, the NWM observations appear unchanging (i.e. little to no variability) over the entire period of record. At Newcomb, NY, neither the Mesonet observations nor the NWM predictions appear to vary much. All three of these sites are located in mountainous regions. Given the variability of soil moisture with topography, one could expect it to be challenging to model soil moisture in mountainous regions. Andes, Old Forge, and Newcomb began collecting data in August, September, and October of 2016, respectively. These poor fits do not

appear to result from potentially erroneous soil moisture observations taken at the beginning of the record (sometimes soil moisture probes require time to equilibrate).

While there appears to be a lack of variability in the 7-day average SM1 soil moisture from the NWM, a number of other things could be impacting the observed results. The poor fits are most likely due to an improper retrieval curve being applied to the soil moisture probes. If the probes were installed with a default curve, one that is not based on the soil conditions at the sites, the observations will inherently be poor. The Mesonet network is a relatively new data source, and to our knowledge little effort has been made to assess the quality of the Mesonet soil moisture readings. As such, unlike the USGS streamflow data used in the previous section, we have less confidence in the soil moisture readings. In general, when the variability of observational data is small, very large negative NSEs may be determined; a small variability in our Mesonet data may be the result of the negative NSEs observed. Soil moisture, which can be impacted by landscape position, can vary in magnitude across small spatial scales. As such the NWM soil moisture may not be capturing conditions at the Mesonet site, especially since our output was obtained from the nearest 1 km grid. It is also possible that simulated soil moisture compared against the Mesonet observations were obtained from a run of the NWM with a soil type not representative of at-site conditions. This could explain the vast differences in the ranges of both the NWM simulated and Mesonet observed soil moisture. Overall, the small NSE and LNSE, and very large RMSE observed here are of concern and indicate either a problem with the NWM soil moisture, the measured soil moisture, the experimental design, or a combination of these factors.

To further examine the poor fits to soil moisture, time series plots for the best (Newcomb), median (Andes), and worst (Old Forge) sites are presented in Figure 13. These figures include observed and simulated soil moisture for the concurrent record (~2016 to present). The observed soil moisture is presented as a red line, while the simulated soil moisture is presented as a black line. At Newcomb, there appears to be relatively good agreement between the observed and predicted soil moisture. The worst

fitting appears as a few periodic events around the spring time. This could mean the influences of rain on snow events may be significantly impacting the quality of the simulated or observed soil moisture. At Andes, there is a systematic over-prediction. However, the general pattern of the observations is captured by the NWM. There appears to be two events, one in the winter of 2016 and another in late fall of 2017, where fitting is particularly poor. In both instances, the observations rapidly decline, a trend the NWM does not capture. Given the inconsistency of the onset of snowpack in northeastern mountainous regions, the influence of early winter snow may be captured by the Mesonet, and not the NWM. At Old Forge, there is an extreme underprediction, and the NWM does not appear to be capturing the temporal extent of the Mesonet in any way. This may be the most prominent example of a site whose retrieval curve does not reflect the at site conditions.

2.4.3 Drought Classification

This analysis now moves to a quantitative and qualitative assessment of the NWM's ability to reproduce drought classifications in NY. As previously stated, county scale data of USDM percentiles was obtained. This data was then compared to the USGS and NWM streamflow data at 19 NY counties, and NWM soil moisture output at the SM1, SM2, and SM3 layers at all 62 NY counties. For the USDM data two separate values were compared against the data in consideration: (1) a weighted average (WA) of the numerical values assigned to all drought categories in a county, and (2) a value assigned based on the category at which the majority of the county (CM) is classified. The bias and RMSE were then calculated for both the WA and CM scenarios. The WA and CM scenarios produced similar results; only the WA scenarios are presented here. Appendix 2 contains the results for the CM scenarios for the 7-day averages, in addition to the WA and CM scenarios for the 14- and 28-day averaging periods. Results were similar, however as the averaging period increased, the bias and RMSE tended to decrease (fitting improved). Figures 14 and 15 present the bias and RMSE of the drought categories for the NWM and USGS streamflow as compared to the USDM drought categories at the county scale. These figures are produced from a comparison of USDM and NWM output (streamflow and soil moisture and USGS observations

(streamflow) over the for the entire period of record of the USDM (January 2000 – present). As expected, the drought categories derived from the USGS streamflow are generally better than those from the NWM streamflow, but results are similar in most counties. This is an encouraging finding, one demonstrating the ability of the various informational sources considered to capture the spatiotemporal extend of the 2016 drought. For the NWM, the bias ranges from 0.26 to 0.94, with a mean of 0.64. This indicates the NWM streamflow (and the USGS streamflow) is more often categorizing a drought of greater intensity than the USDM. One important issue is with the period of record considered. While the NWM is constrained to the reanalysis period in this analysis (and the USGS streamflow is constrained over a similar concurrent period), the USDM employs a longer historic record when developing its drought categories. It is possible that our percentiles are artificially inflated due to the relatively short period of record as they may be more sensitive to recent departures from the climatological normal. Results for the RMSE in Figure 15 indicate a relatively large RMSE compared to the magnitude of the drought classes; this indicates that in some cases the NWM and USGS streamflow categories are classified as more than one drought category away from the USDM.

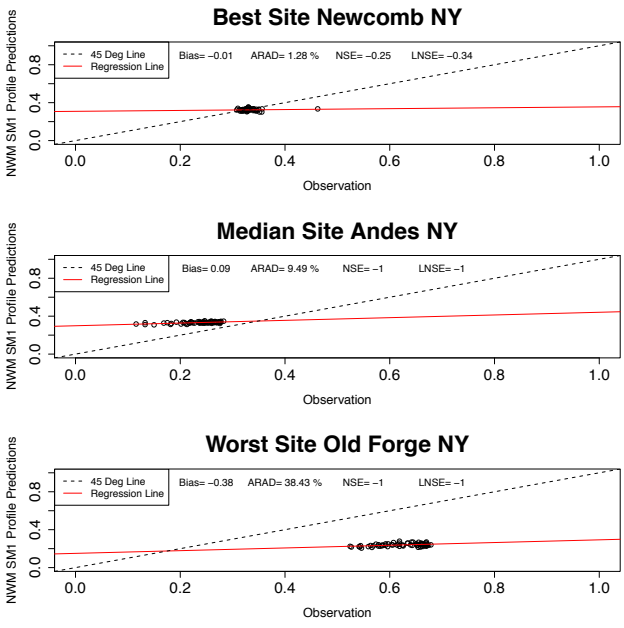


Figure 12- Plots of NWM 7-day average SM1 profile soil moisture versus Mesonet observed 7-day average soil moisture at site with the lowest ARAD, median ARAD, and highest ARAD.

With RMSE, the USGS streamflow is generally better than the NWM soil moisture, but again the USGS and NWM streamflow categories are similar in performance.

Figure 16 presents the bias from the comparison of NWM soil moisture at all three soil moisture layers to the county USDM weighted average categories. In the SM1 layer, the bias has a range from -0.08 to 1.1, with a

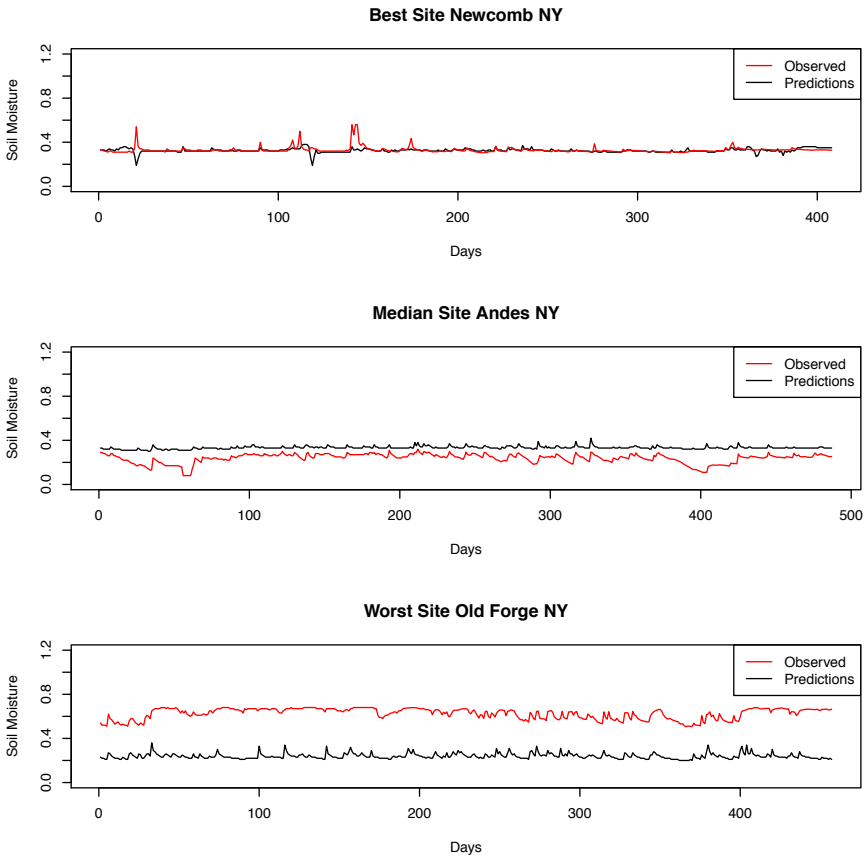


Figure 13- Plot of Time Series Soil Moisture at the Best site (Newcomb NY), Median Site (Andes NY), and Worst Site (Old Forge NY) Identified from the 7-Day Average AAD

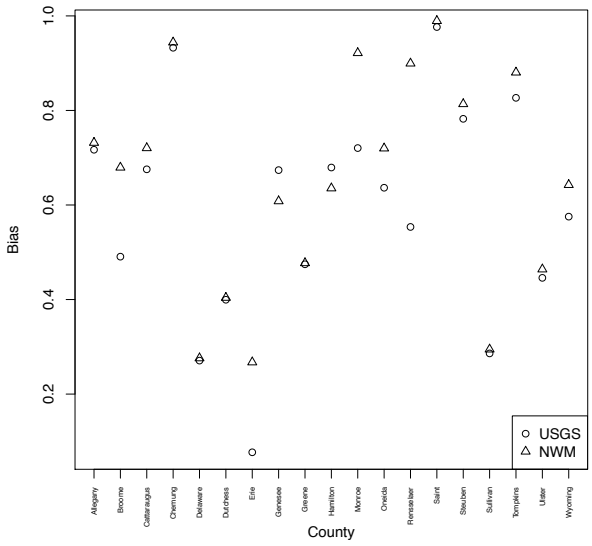


Figure 14- Bias for 7-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

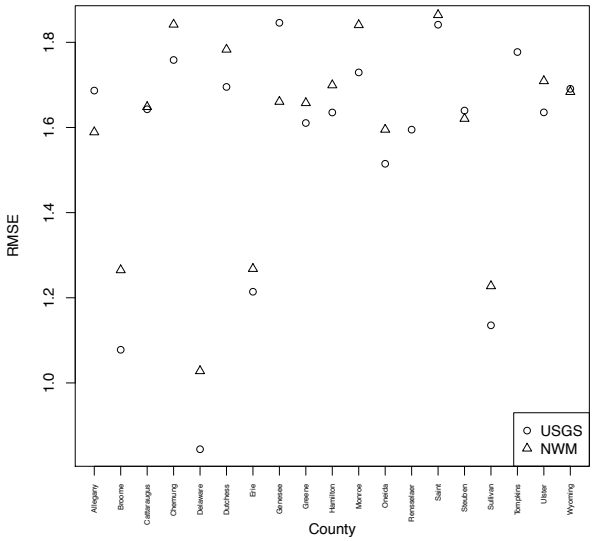


Figure 15- RMSE for 7-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

mean value of 0.52. Results for SM2 and SM3 layers were similar. For all three soil moisture layers, it appears that the NWM is again overpredicting the extent of drought in NY relative to that of the USDM. This may seem contradictory to the general overestimation of streamflow or soil moisture determined from the numerical assessment. The percentiles here were derived based on the flow values. The 30th

percentile is an order statistic which does not depends on the other flow values being considered, and is not necessarily overestimated at sites where streamflow or soil moisture is overestimated. In general, the observed drought category bias for the NWM soil moisture is less than the observed drought category bias for the NWM streamflow, indicating the soil moisture is potentially better at characterizing drought categories than streamflow, consistent with the results found by Xia et al. (2014). The RMSE for NWM soil moisture in Figure 16 appear large, especially in relationship to the magnitude of the bias. For all three profiles, the median bias was 0.45, while the median RMSE, for all profiles, was 1.37, reflecting the strong impact that a misclassified county may be having on the analysis. Since, in this experimental design, drought categories work on a step (i.e. -1 for ND, 0 for D0 etc.), a county misclassified by more than 1 category (i.e. D2 vs. D0) may have a relatively large impact on the bias, but this will only be magnified with the RMSE. As previously discussed, the overestimation of the drought categories by the NWM soil moisture may be due to the different records employed compared to the USDM.

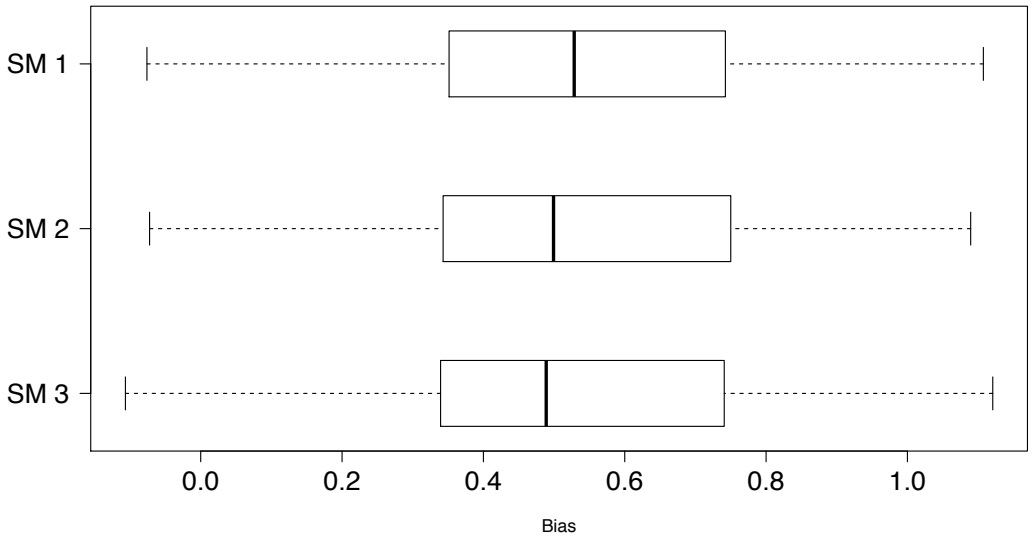


Figure 16- Bias for 7-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories.

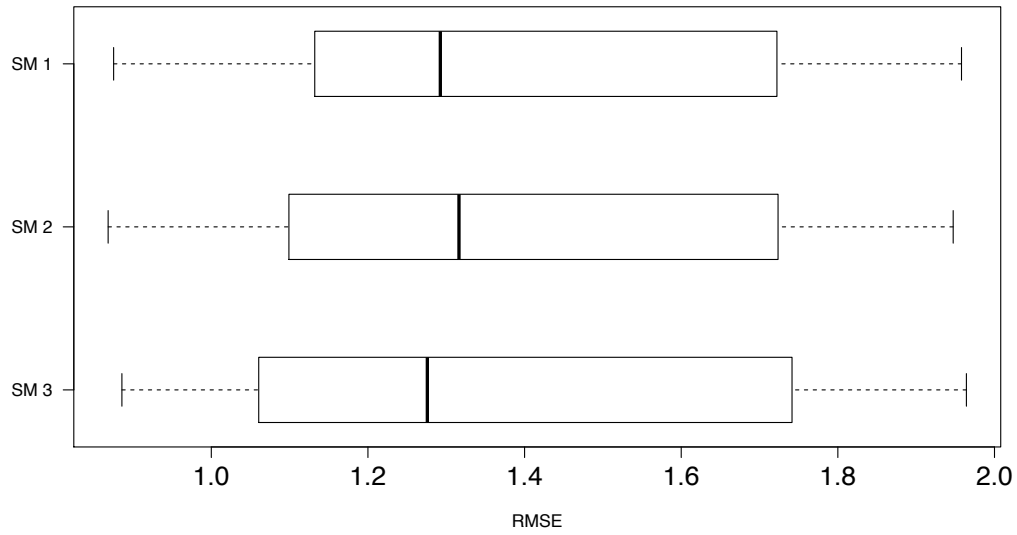


Figure 17- RMSE for 7-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories.

As a final analysis, a qualitative visual assessment of drought categories is presented. Figure 18 contain the USDM map for September 6, 2016, at the height of the 2016 northeastern U.S. drought. One observes that about 80% of NY was in a drought category of D0 or higher, and about 20% of NY was in D3 (extreme drought). Also, on Figure 18 are the drought categories from the NWM SM1, SM2 and SM3 soil moisture profiles for the Mesonet locations on September 6 (based on 7-day average soil moisture), and drought categories from the NWM and USGS streamflows. In general, it appears the NWM is able to capture the general spatial extent of the 2016 drought. The USDM places D3 designations over much of western NY. This is generally captured by both the streamflow (bottom left and right of Figure 18) and soil moisture classifications (top right, and middle left and right of Figure 18). The USGS streamflow appears to be under classifying the drought with respect to the USDM boundaries, a sharp contrast to the NWM output (either streamflow or soil moisture). The SM3 profile appears to emulate the USDM boundaries most accurately. In northwestern NY, in the SM1 and SM2 profiles, many of the sites are classified by the NWM as ND, which is does not align well with the boundaries determined by the USDM. The SM3 profile seems to better capture these boundaries, classifying northwestern NY more consistently with the USDM maps.

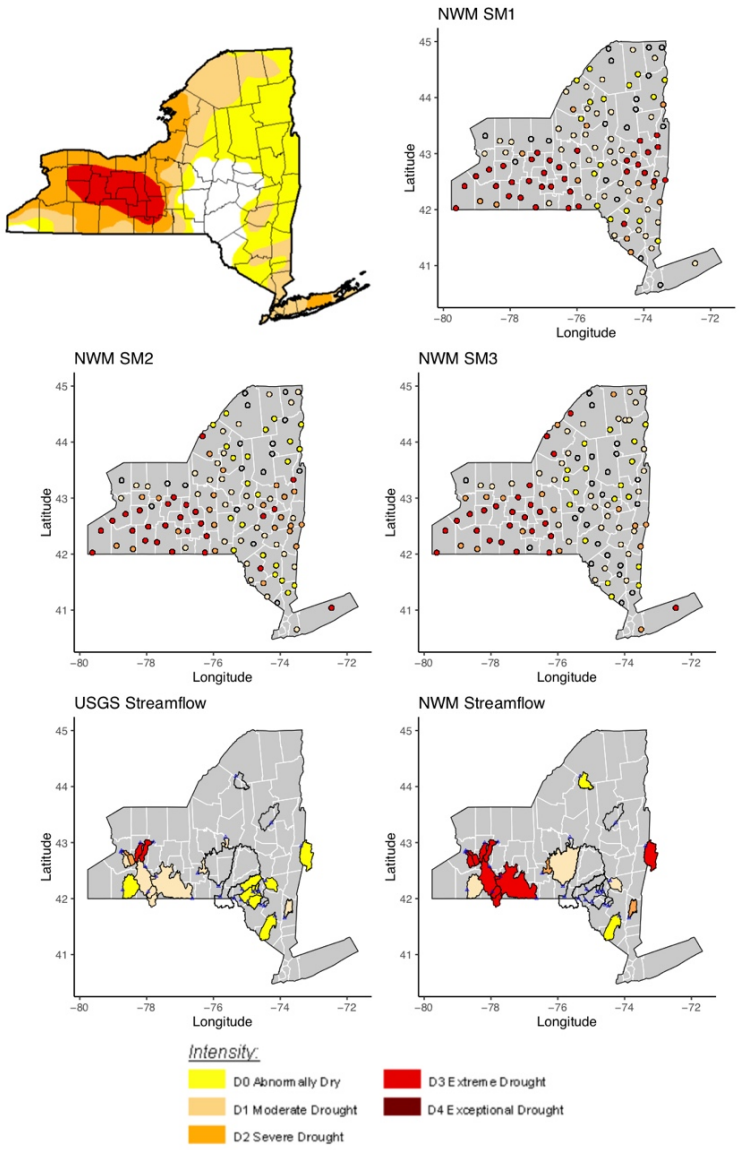


Figure 18- September 6, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

The NWM streamflow map also appears to generally capture the spatial extent of the drought, while the USGS streamflow map actually appears to underclassify the drought, contrasting the general over classification of drought associated with NWM output (streamflow or soil moisture).

One notable exception is in eastern NY (northern portion of the Hudson Valley region) where the NWM streamflow and soil moisture data generally predicted D3, while the USDM generally predicted D1 or D2. The NWM soil moisture data in all three layers seems to reflect a more extreme drought in this region than does the USDM, representing the most substantial difference between our derived drought categories and those of the USDM. The misclassification of drought in this region of NY may result from the different periods of record considered in the USDM as opposed to the NWM classifications. It is also possible that the USDM in this region of NYS bereft of stakeholder inputs. Currently, the NWS has 3 local forecast offices in NY (Buffalo, Albany and Binghamton). At or around these locations, one would expect the most accurate USDM boundaries. It is somewhat troubling that in the vicinity of the Albany station a severe mismatch is observed. However, it is possible that given the lack of population just east of the Albany area, the Upper Hudson Valley region may be lacking on-the-ground reports, therefore decreasing the accuracy of the USDM in these areas.

Additional figures for the first USDM map released in each month from May 2016 – April 2017 are presented in Appendix 2. This period captures the onset, peak, and decline of the 2016 drought in New York State. The maps generated with the informational sources presented above appeared to do similarly well for the entirety of the 2016 drought. This is an encouraging finding, as it appears the NWM reproduces drought categories in NY with reasonable accuracy regardless of season.

2.5 Summary and Conclusions

This analysis presented a characterization of drought in New York considering various data from the National Water Model (NWM). A numerical comparison of streamflow and soil moisture data was conducted first. NWM streamflow from a 25-year reanalysis period was compared to USGS streamflow, and various performance metrics were calculated for 7-, 14-, and 28-day average streamflow at 28 unregulated gauging stations in NY. Metrics werecalculated for the entire record and for x-day averages below the 30th percentile.

In general, the NWM did reasonably well at predicting streamflow in NY. It was observed that for longer averaging periods (14-, and 28-day), the accuracy of the NWM generally increased. In general, the NWM performs similarly, yet slightly worse for the smaller flows, potentially much worse at smaller basins. While at some sites the absolute error of the NWM 7-day average streamflows was as low as 30% on average, in general the NWM had a positive bias (overestimation) and an absolute error of more than 50% at over half of the stream reaches considered. While the absolute error was large several sites, in general the NWM appears to only do moderately well at simulating observed USGS streamflows (i.e. results were similar to Xia et. al (2014) who employed NLDAS-2 in a similar experiment).

Numerical comparisons were then made between NWM and New York Mesonet soil moisture data. Identical averaging periods, and similar non-relative statistics were calculated Fitting to the soil moisture was determined to be poor, and may be attributed to either the Mesonet observations, or the NWM output. A major issue with the Mesonet data is its short period of record, one that is not concurrent with the reanalysis period used in this analysis. In an attempt offset this issue, three longer term soil moisture sites (Geneva SCAN, Ithaca CRN, and Millbrook CRN) were used to compare against the nearest site at which NWM soil moisture data was available. Similar to streamflow, as the averaging period increased, NWM generally provided better agreement with the observations. This may be attributed to a reduction in the

variance as the averaging period increased. Further, the deeper soil moisture layers (SM2 and SM3) were better represented by the NWM. This is not surprising due to the influence topography and rapidly changing near-surface environmental conditions may have on shallow soil moisture (SM1) corresponding to a reduction in the variance naturally associated deeper soil moisture observations. It is generally more difficult to model soil moisture at the surface layers (Manfreda et al. 2007). However, overall, the NWM soil moisture data was a poor predictor of Mesonet soil moisture. At more than 50% of the sites in each SM layer, the NSE was less than -1. Given that an NSE value 0 represents the point where simply taking the average of the observations would provide a better estimation of soil moisture than would NWM simulation, NSEs below -1 are concerning. After inspecting several plots of soil moisture, it was determined that the NWM 7-day average soil moisture output lacked much variability over the periods examined relative to the Mesonet data. The Mesonet is a new relatively new data source, and it is possible some problems may be present in this data set. There may also be issues with the NWM in this region. The NWM similarly lacks vetting in the Northeast, and applies a fixed seasonal vegetation cycle, which does not respond well to flooding and drought.

The development of drought categories in NY was conducted next. USGS streamflow, and NWM streamflow and soil moisture data were used to derive drought categories based on the percentile ranges of non-exceedance probabilities established by the USDM. Weekly county-scale drought classifications were obtained from the USDM and these were numerically compared to drought categories from the USGS streamflow and NWM streamflow and soil moisture series over the USDM period of record (January 2000 – present). The drought category bias was generally positive for both USGS streamflow and NWM output (streamflow and all soil moisture layers), indicating that our data generally predicted a more intense drought than the USDM. This overestimation of drought may be due to the base period of the USDM dating back to 1979 for some data. While the USGS and NWM data was analyzed only over the reanalysis period (1993-2017), the USDM is partly based on longer-term historical data, although most of the remote sensing data is only available from 2000-present. If the period from 1993-2017 had

fewer drought periods than the USDM historical record, one would generally overestimate the intensity of droughts during the reanalysis period relative to that of the USDM datasets. Finally, a visual assessment of the spatial extent of drought in NY in September 2016 during the height of the 2016 drought was performed. In general, the NWM appeared to do a relatively good job capturing the spatial extent of the 2016 drought, especially in the SM3 soil moisture layer (the deepest considered) doing the best to reproduce the USDM map.

While results were somewhat mixed, overall this analysis showed the NWM output has some skill in reproducing drought series in NY. The version of the NWM used to create the reanalysis output has a relatively simplistic groundwater model, but our observations that drought categories based on NWM streamflow were similar to those based on USGS streamflow is promising. Further exploration of how best to use Mesonet soil moisture to inform the NWM model is also warranted, as deeper soil moisture appears to have the best skill in characterizing drought in NY. Further exploration should also be given to the USDM, as to the best of our knowledge, a critical evaluation of the USDM, and its data products has not been undertaken in the Northeast. The USDM has a similar set of limitations which have been detailed to some extent in this work. Comparing against the USDM is advantageous however, given its use to generate federal disaster aid for farmers in the US.

2.6 References

Bengtsson, L., Hagemann, S., & Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, *109*(D11).

Bradford, N. (2018). The Increasing Demand and Deceasing Supply of Water. National Environmental Education Foundation of the United States of America (NEEF USA*)*. Retrieved January 18, 2019, from https://www.neefusa.org/nature/water/increasing-demand-and-decreasing-supply-water.

Donnelly, K., & Cooley, H. (2015). Water Use Trends. *Oakland, California: Pacific Institute*.

Farmer, W.H., & Vogel, R.M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research,* https://doi.org/10.1002/2016WR019129

Flanders, A., E. Bauske, and J. McKissick, 2008: Economic impact of total watering restrictions to the green industry in the Bear Creek Reservoir region. Center for Agribusiness and Eco- nomic Development Rep. CR-08-05, University of Georgia, 8 pp. [Available online at http://www.caes.uga.edu/center/caed/ pubs/2008/documents/CR-08-05.pdf.]

Garen, D. C. (1993). Revised surface-water supply index for western United States. *Journal of Water Resources Planning and Management*, *119*(4), 437-454.

Gochis, D.J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J. Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, W. Yu, (2018). The WRF-Hydro modeling system technical description, (Version 5.0). NCAR Technical Note. 107 pages. Available online at https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf. Source Code DOI:10.5065/D6J38RBJ

Gochis, D., A. Dugger, J. McCreight, L. Karsten, W. Yu, L. Pan, D. Yates, Y. Zhang, K. Sampson, B. Cosgrove, F. Salas, E. Clark, T. Graziano, D. Maidment, C. Phan, Z. Cui, Y. Liu, X. Feng, H. Lee. (2019). "Technical Description of the National Water Model Implementation WRF-Hydro." National Center for Atmospheric Research (NCAR). Retrieve on June 1, 2019, from https://www.cuahsi.org/uploads/cyberseminars/wrf_hydro_nwm_cuahsi_nwm_webinar_oct_2016_%281%29.pdf.

Guerrero, B. (2012). The impact of agricultural drought losses on the Texas economy, 2011. *Briefing Paper, AgriLife Extension*.

Hayhoe, K., Wake, C.P., Huntington, T.G., Luo, L., Schwartz, M.D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J., DeGaetano, A. and Troy, T.J., 2007. Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dynamics*, *28*(4), pp.381-407.

Howitt, R., Medellín-Azuara, J., MacEwan, D., Lund, J. R., & Sumner, D. (2014). *Economic analysis of the 2014 drought for California agriculture*. University of California, Davis, CA: Center for Watershed Sciences.

Kohl, E., & Knox, J. A. (2016). My drought is different from your drought: A case study of the policy implications of multiple ways of knowing drought. *Weather, Climate, and Society*, *8*(4), 373-388.

Manfreda, S., McCabe, M. F., Fiorentino, M., Rodríguez-Iturbe, I., & Wood, E. F. (2007). Scaling characteristics of spatial patterns of soil moisture from distributed modelling. *Advances in water resources*, *30*(10), 2145-2150.

Mo, K. C., & Lettenmaier, D.P. (2016). Precipitation Deficit Flash Droughts over the United States. *J. Hydrometeor.*, 17, 1169–1184, https://doi.org/10.1175/JHM-D-15-0158.1.

Nash, J. E.; Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part I — A discussion of principles". *Journal of Hydrology*. **10** (3): 282–290. doi:10.1016/0022-1694(70)90255-6.

Nielsen-Gammon, J. (2011). *The 2011 Texas drought: a briefing packet for the Texas Legislature*.

National Center for Atmospheric Research University Cooperation for Atmospheric Research Applications Laboratory Weather Research and Forecasting Hydrologic Model (2019). WRF-Hydro Modeling System. Retrieved on June 30, 2019, from https://ral.ucar.edu/projects/wrf_hydro/overview.

National Oceanic and Atmospheric Administration National Center for Environmental Protection (2019a). The NCEP Climate Forecasting System Version 2. Retrieved June 20, 2019 from https://cfs.ncep.noaa.gov.

National Oceanic and Atmospheric Administration (2019b). Center for Satellite Applications Research Global Vegetation Health Products. NOAA STAR Homepage. Retrieved March 4, 2019, from https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_browseVH.php.

National Oceanic and Atmospheric Administration Office of Water Protection (2019c). Information About the National Water Model. Retrieved June 20, 2019, from http://water.noaa.gov/about/nwm.

National Oceanic and Atmospheric Administration Earth System Research Laboratory (2019d). Global Forecasting System. Retrieved on June 30, 2019, from https://www.emc.ncep.noaa.gov/GFS/.php.

National Oceanic and Atmospheric Administration Earth System Research Laboratory (2019e). The High Resolution Rapid Refresh. Retrieved on June 30, 2019, from https://rapidrefresh.noaa.gov/hrrr/.

National Oceanic and Atmospheric Administration Environmental Modeling Center (2019f). National Stage IV QPE produce. Retrieved June 30, 2019, from https://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/.

National Oceanic and Atmospheric Administration National Climactic Data Center (2019g). United States Climate Reference Network. Retrieved June 15, 2019, from https://www.ncdc.noaa.gov/crn/qcdatasets.html.

National Oceanic and Atmospheric Administration The National Severe Storms Laboratory (2019h). Multi-radar/Multi-sensor System (MMRS). Retrieved June 30, 2019, from https://www.nssl.noaa.gov/projects/mrms/.

New York State Mesonet (2019). About New York's Mesoscale Weather Network. Retrieved June 30, 2019, from http://www.nysmesonet.org/#about.

Northeast Drought Early Warning System (DEWS) partners (2018). Northeast Drought Early Warning System 2018-2019 Strategic Plan. Prepared by the National Integrated Drought Information System (NIDIS) in partnership with key stakeholders including the Northeast Regional Climate Center (NRCC) and NOAA's Regional Climate Services Director, Eastern Region.

Northeast Regional Climate Center (NRCC), Cornell University (2016). Coping with Drought and its Aftermath in the Northeast. Retrieved on March 4, 2019, from http://www.nrcc.cornell.edu/regional/drought/pubs/assessment_2016.pdf

Ogden, F.L. (1997). CASC2D Reference Manual. Dept. of Civil and Evniron. Eng. U-37, U. Connecticut, 106 pp.

Otkin, J. A., Svoboda, M., Hunt, E. D., Ford, T. W., Anderson, M. C., Hain, C., & Basara, J. B. (2018). Flash Droughts: A Review and Assessment of the Challenges Imposed by Rapid-Onset Droughts in the United States. *Bull. Amer. Meteor. Soc.*, 99, 911–919, https://doi.org/10.1175/BAMS-D-17-0149.1.

Palmer, W.C. (1965). Meteorologic Drought. US Department of Commerce, Weather Bureau, Research Paper No. 45, p. 58.

Postel, S. L., Daily, G. C., & Ehrlich, P. R. (1996). Human appropriation of renewable fresh water. *Science*, *271*(5250), 785-788.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., and Gochis, D. J. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions, *Bulletin of the American Meteorological Society,* https://doi.org/10.1175/BAMS-D-15-00308.1.

Rippey, B. R. (2015). The US drought of 2012. *Weather and Climate Extremes*, *10*, 57-64.

Read, L., Gochis, D., Dugger, A., Yates, D., Fitzgerald, K., Sampson, K., & Salas, F. (2018, April). Overview of Development and Evaluation of the National Water Model. In *EGU General Assembly Conference Abstracts* (Vol. 20, p. 11694).

Seager, R., Pederson, N., Kushnir, Y., Nakamura, J., & Jurburg, S. (2012). The 1960s drought and the subsequent shift to a wetter climate in the Catskill Mountains region of the New York City watershed. *Journal of Climate*, *25*(19), 6721-6742.

Shiklomanov, I. A. (1998). World water resources. *A new appraisal and assessment for the 21st century*.

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., & Vogt, J. (2014). World drought frequency, duration, and severity for 1951–2010. *International Journal of Climatology*, *34*(8), 2792-2804.

Stedinger, J.R. (1981). Estimating correlations in multivariate streamflow models, *Water Resources Research,* 200 – 208.

Stedinger, J.R., Vogel R.M., Foufoula-Georgiou E. (1993). *Handbook of hydrology* (Ch. 18 Frequency Analysis of Extreme Events). New York: McGraw-Hill.

Sullivan, K. (2016). National Water Model Improving NOAA's Water Prediction Services. NOAA. Retrieved January 22, 2019, from http://water.noaa.gov/documents/wrn-national-water-model.pdf

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D. and Miskus, D., 2002. The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), pp.1181-1190.

Sweet, S. K., Wolfe, D. W., DeGaetano, A., & Benner, R. (2017). Anatomy of the 2016 drought in the Northeastern United States: implications for agriculture and water resources in humid climates. *Agricultural and Forest Meteorology*, *247*, 571-581.

United Nations (UN). (2019). Land and Drought. Accessed at https://www.unccd.int/issues/land-and-drought on July 5, 2019.

United States Department of Agriculture Natural Resources Conservation Service. (2019). Soil Climate Analysis Center Data and Products. Retrieved June 15, 2019, from https://www.wcc.nrcs.usda.gov/scan/.

United States Drought Monitor (USDM) (2019). United States Drought Monitor: What is the U.S. Drought Monitor. Retrieved on June 15, 2019, from https://droughtmonitor.unl.edu/AboutUSDM/WhatIsTheUSDM.aspx

United States Geologic Survey. (2018). WaterWatch: Drought. Accessed June 6, 2019, from https://waterwatch.usgs.gov/index.php?id=ww_drought

Varol, T., & Ertuğrul, M. (2016). Analysis of the forest fires in the Antalya region of Turkey using the Keetch–Byram drought index. *Journal of forestry research*, *27*(4), 811-819.

Vieux, B. E. (2001). Distributed hydrologic modeling using GIS. In *Distributed Hydrologic Modeling Using GIS* (pp. 1-17). Springer, Dordrecht.

Weibull, W. (1939). A statistical theory of strength of materials. *Ing. Vetensk. Akad. Handl.,* 151, 1–45.

Xia, Y., Ek, M. B., Peters-Lidard, C. D., Mocko, D., Svoboda, M., Sheffield, J., & Wood, E. F. (2014). Application of USDM statistics in NLDAS-2: Optimal blended NLDAS drought index over the continental United States. *Journal of Geophysical Research: Atmospheres*, *119*(6), 2947-2965.Donnelly, K., & Cooley, H. (2015). Water Use Trends. *Oakland, California: Pacific Institute*.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J. and Livneh, B., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

Chapter 3: Conclusion

3.1 Thesis Summary

Accurately characterizing the spatiotemporal extent and intensity of drought is critical to water resource management and disaster relief recovery. The USDA Farm Service Agency (FSA) currently employs the U.S. Drought Monitor (USDM) drought categories to base its allocation of emergency relief aid to farmers affected by drought (FSA 2012). Using a wide variety of information and indices, the USDM classes drought in categories: ND, no drought; D0, abnormally dry; D1, moderate drought; D2, severe drought; D3, extreme drought; and D4, exceptional drought (USDM 2019). Livestock owners in counties classified as D2 or greater for at least 8 consecutive weeks are eligible for a single monthly payment. As drought increases in severity, the payments increase accordingly (e.g. livestock owners in counties with a D4 at any time, or D3 for 4 consecutive weeks are eligible for 3 monthly payments) (FSA 2019). Crop farmers receive relief by a slightly different process. A USDA FSA representative will designate a county as disaster if a county is under D2 for 8 consecutive weeks, or D3 at any point during the growing season. Once a county is designated as a disaster area, emergency funding will be made to eligible recipients in the county (FSA 2012). However, there are provisions such that when farmers in a county report > 30% crop production losses, a county may be receive a disaster designation at which point emergency funding would be made available (FSA 2012).

This thesis focused primarily on exploring new data sources to augment the USDM. The primary data source was a 25-year reanalysis output from NOAA's National Water Model, an advanced hydrologic modelling framework that simulates and forecasts streamflow over the continental U.S (Read et al. 2018). This thesis first performs a numerical assessment of the NWM's reanalysis streamflow output to predict USGS gauged streamflow. Next a numerical assessment is performed on how the NWM's reanalysis soil moisture output compares to soil moisture measured at New York Mesonet sites at three different depths in the soil column (SM1, 0 – 10 cm; SM2, 10 – 40 cm; SM3, 40 – 100 cm). Due to the short concurrent

record of the soil moisture observations, NWM soil moisture was also compared to observation at three sites with longer-term records.

With the numerical comparison completed, drought categories were derived for the USGS and NWM streamflow, and NWM soil moisture output. Derived drought categories were then compared to USDM drought categories over the entire period of the USDM. Finally, as a visual means of comparison, drought maps were developed employing derived drought categories for streamflow and soil moisture and compared to the USDM map for New York on September 6, 2016, at the height of the 2016 Northeastern U.S. drought.

The major findings of this thesis are as follows

1) NWM streamflow had varying degrees of agreement with USGS observed streamflow, with average absolute error ranging from 30% to over 100%, and generally a positive bias. While prediction appears to be slightly worse for the smaller streamflows at a specific site, in general the worst prediction appeared to be at smaller watersheds.
2) NWM soil moisture was generally not a good predictor of observed soil moisture (at both Mesonet and long term sites). An overall positive bias was determined, indicating a general overprediction of soil moisture. It should be noted that the Mesonet and NWM reanalysis soil moisture values are from relatively new data sources which may require additional vetting. In general, the NWM soil moisture did not seem to vary temporally as one would expect.
3) NWM output seemed to capture the general spatial patterns of the northeastern drought of 2016 in NY. The NWM soil moisture and streamflow data generally emulated the USDM map, with deeper soil moisture providing a better agreement with the USDM.

3.2 Reflection on Research Objectives

Chapter 1 presented two research goals which formed the basis of this research. Both will be individually revisited with an accompanying brief discussion about the degree in which these goals were met.

Research Goal 1: Examine how NWM streamflow and soil moisture output compares to observed streamflow and soil moisture observations during drought periods.

A relatively robust analysis of the data was explored. The relative bias (RBias), average relative absolute deviation (ARAD), Nash-Sutcliffe Efficiency (NSE) in real- and log-space and the relative root mean square error (RRMSE) were calculated for streamflow comparisons, and bias, AAD, NSE and RMSE for soil moisture comparisons for three different averaging periods (7-, 14-, and 28-day). This research was a successful initial analysis that identified potential issues with the data and assessed the NWM's ability to reproduce streamflow and soil moisture observations.

Research Goal 2: Explore how the NWM characterizes the intensity of drought in NY and to compare this to drought intensity from the USDM, with a particular focus on the drought of 2016.

Drought categories were derived using percentiles obtained from the non-exceedance probabilities for USGS and NWM streamflow data and NWM soil moisture data. These percentiles were then numerically compared to county-scale USDM data. The bias and RMSE were determined for each county, and maps were produced as a visual means of comparison. It was observed that the NWM was able to capture the general spatial extent of the 2016 drought in NY.

3.3 Future Direction

This thesis provides an initial analysis of the use of NWM output to characterize drought in NY. There are many potential steps forward which include:

1) Expand analysis to encompass the entire northeast. The drought of 2016 impacted the northeast region as a whole, so it makes sense to expand this analysis to a wider region. This would allow for an assessment of NWM output at more streamflow and soil moisture sites.

2) Consider the influence groundwater has on drought. We made no determination of the groundwater lost during streamflow recessions or the resulting groundwater levels during drought. This information could help improve water supply management during drought.
3) Consider a more encompassing subset of streamflow gauges in NY to better understand the impact of streamflow on drought characterization across a wider region. We considered streamflow in less than 50% of NY counties, and nearly 20% of our sites were in one county.
4) Further investigate the NWM and its ability to model streamflow in smaller watersheds.
5) Further investigate NY Mesonet soil moisture data. Results indicate a poor fit of NWM soil moisture to Mesonet soil moisture, indicating a potential issue with one or both of these data sources.
6) Evaluate any seasonal biases in the soil moisture. It is possible that the NWM/Mesonet are predicting extremely poor for a few localized events which heavily influence the poor numerical values of the statistics.
7) More fully investigate how the USDM incorporates information to produce the weekly maps. The USDM may not be the best metric to compare against given the varied spatial scale associated with many of its inputs.
8) Evaluate the USDM as ”truth” to be compared against. The USDM is not without its limitations ant it may not be the best metric for which to assess multi-model output, and its ability to predict drought in NY.
9) Access digital USDM maps. This analysis used county specific drought information. Should digital USDM maps become available, they may help us more fully understand how the NWM output reproduces USDM categories.
10) Explore the use of the deeper NWM soil moisture output (SM4), which was not employed in this analysis. The deepest soil moisture profile examined (SM3) appeared to best match the USDM drought categories. This deeper probe may provide additional regional drought information.

3.4 References

Read, L., Gochis, D., Dugger, A., Yates, D., Fitzgerald, K., Sampson, K., & Salas, F. (2018, April). Overview of Development and Evaluation of the National Water Model. In *EGU General Assembly Conference Abstracts* (Vol. 20, p. 11694).

Farm Service Agency (FSA), United States Department of Agriculture (2012). Federal Register/Vol. 77, No. 135/Friday, July 13, 2012/Rules and Regulations. 7CFR Part 1945. RIN 0560-AH17. Disaster Designation Process. Retrieved on July 4, 2019, from https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/Disaster-Assist/Secretarials/Presidential-Metadata/FR_2012-17137.pdf

Farm Service Agency (FSA), United States Department of Agriculture (2019). Livestock Forage Disaster Program Fact Sheet March 2019. Retrieved July 4, 2019, from https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/FactSheets/2019/livestock_forage_disaster_program.pdf

United States Drought Monitor (USDM) (2019). United States Drought Monitor: What is the U.S. Drought Monitor. Retrieved on June 15, 2019, from https://droughtmonitor.unl.edu/AboutUSDM/WhatIsTheUSDM.aspx

Appendix 1: An Assessment of i-Tree Hydro

A1.Abstract

i-Tree Hydro is a tool for modeling the impact of land cover changes on hydrology and hydrologic processes, with a particular focus on the tradeoffs between trees and impervious surfaces. The ensuing report details an assessment of i-Tree Hydro and provides a list of suggested improvements to its developers (Appendix A1.A). Issues involving model input, processes, calibration and output are investigated. A sensitivity analysis was performed to identify an improved subset of model parameters for calibration. Additionally, an alternative calibration objective function based on monthly runoff ratios was explored with the goal of improving how i-Tree Hydro represents the impact of trees on the urban water balance. Finally, numerous potential issues with the model were identified. The i-Tree Hydro development team should consider these issues as they may improve the use of this model. Note that for the entirety of this assessment, i-Tree Hydro version 6.1.1 was used.

A1.1.0 Introduction

Trees often play an integral role in the hydrologic cycle. Trees have the ability to intercept water, aid in evapotranspiration, and reduce and improve stormwater quantity and quality, with an additional benefit of cooling the local environment which is particularly important in urban settings. i-Tree Hydro has many routines specifically tailored to capture the effects trees have on the quantity and quality of runoff in urban watersheds. It was developed to help users determine how land cover changes may influence the hydrologic cycle, especially in urban and semi-urban settings (which we will refer to generally as “urban” in the rest of this document).

One problematic scenario for most urban areas is a “flashy” response to precipitation events. Urban areas typically include compacted, non-native soils, and the addition of infrastructure and drainage systems that alter the watershed’s hydrologic response. Water generally requires more time to infiltrate compacted soils than soils under natural conditions. Additionally, the placement of roads, parking lots, buildings, and

drainage systems all generally increase the quantity and velocity of surface runoff. Considering there is far less resistance to flow, and a decreased ability for water to infiltrate the soil layers, it is not surprising to observe the flashy responses urban watershed often demonstrate during precipitation events.

Fortunately, trees are an effective means to mitigate the flashy response of urban watersheds to precipitation events. By planting trees, the urban watershed more closely resembles its pre-development conditions. Tree root systems can increase infiltration by loosening the soil structure. Their leaves and stems can also intercept water, decreasing the total volume of water that reaches the surface. Trees generally promote evapotranspiration, a net removal of water through the atmosphere, and thus can directly contribute to a reduced volume of surface runoff. These processes can also dampen flood waves that may have otherwise been detrimental to the urban landscape.

Clearly, it is advantageous for city planners and managers to have a tool which captures the role trees play in the urban water cycle; i-Tree Hydro (Hydro) is one such tool. The ensuing report presents findings on Hydro and provides a description of potential problems with, and ways to improve, the model. The following sections discuss the aggregation of streamflow and the study sites, a model sensitivity analysis, alternative calibration scenarios, and suggestions for model improvement.

A1.2.0 Streamflow Aggregation and Study Sites

The initial goal of this study was to replicate the Nash-Sutcliffe Efficiency (NSE) from a calibration on the weekly residual sum of squares (RSS) using the statistical computing program R. The intended outcome of this was to ensure that Hydro was handling this calculation correctly, to familiarize the authors with Hydro code, and to develop a calibration routine outside of the Hydro modeling structure to explore alternative calibration approaches. The NSE is calculated by:

$$NSE = 1 - \left(\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^{N}(\theta_i - \bar{\theta}_i)^2}\right) \quad \text{(A1.2)}$$

where $\widehat{\theta_i}$ is the model prediction at time i, θ_i is the observation at time i, N is the number of time steps, and $\overline{\theta_i}$ is the average of the observations over the N time steps. The Hydro Graphical User Interface (GUI) produces an output file, qobs.dat, that was used to perform a manual aggregation of the Hydro processed observed streamflow data to independently replicate the NSE produced by Hydro. Hydro requires sub-hourly streamflow data as an input, which is then processed to an hourly time step. The aggregation is completed by taking the first observation of any given hour and using that value as the hourly average streamflow. The hourly data was then manually aggregated to a desired time step (daily, weekly, or monthly) for further analysis. The United States Geologic Survey (USGS) determines daily discharge values from instantaneous data (or sub-hourly data, whichever is available) by taking the average of the discharges over an entire day (USGS 2018). We encourage the i-Tree Hydro development team to work towards updating Hydro to reflect how the USGS handles its daily aggregation. It is entirely possible that within an hour, a high intensity, short duration precipitation event may affect observed discharge, especially in smaller urban watersheds. Given that Hydro is only considering the first observation of any hour, it may be misrepresenting the impact of such storms. Although in many situations, the impact of this aggregation will be minimal. Figures A1.1, A1.2, and A1.3 show the differences between aggregating 5-minute USGS historical streamflow observations, 15-minute USGS historical streamflow observations, and the first hourly reading (Hydro's aggregation), all to the daily time step, at Dead Run in Franklintown, MD (USGS Gauge #01589330) in June 2007. The main differences observed are during precipitation events when the streamflow is rapidly changing.

Further, if no observation exists within a given hour, Hydro will linearly interpolate a streamflow observation for that hour. The USGS handles such a scenario by acquiring data from a near-by gauging station or via inference from a precipitation record (USGS 2018). We encourage the i-Tree Hydro development team to consider calibrating only on time steps with recorded data if the calibration routine continues to be based off sub-hourly or instantaneous data, and avoid using interpolated, or inferred data. Another suggestion would be to update Hydro to require USGS daily data as a model input rather than

sub-hourly or instantaneous data, which is the current requirement. This should help alleviate the issue of calibrating on inferred data and would allow the model to employ a vetted methodology (via USGS protocols) for obtaining observed streamflows. One potential drawback to such an update is that this does not allow the user to calibrate on sub-daily streamflows, but such calibrations should probably be avoided so that the model better represents the overall water balance rather than trying to fit short-term peak streamflows.

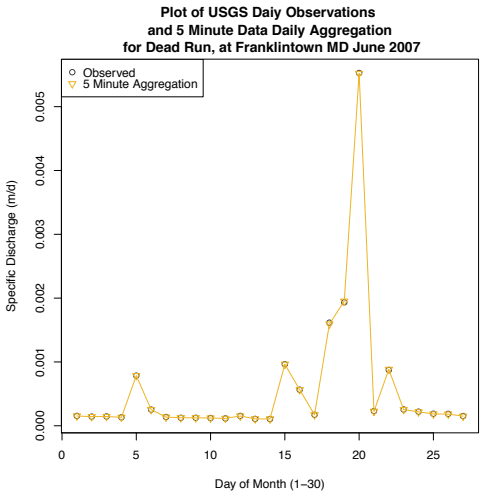


Figure A1.1- Plot of USGS Surface Water Daily Data, and a daily aggregation of USGS Surface Water Historical Observations (5-minute data).

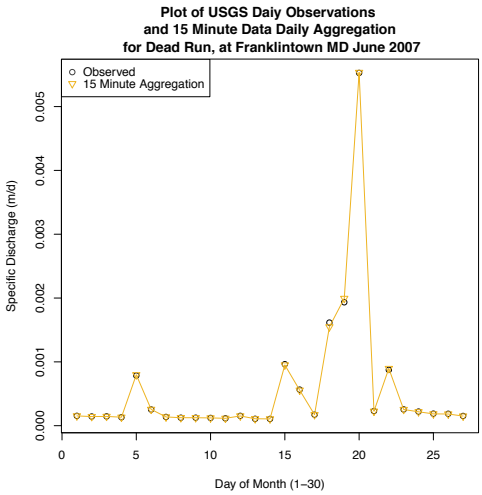


Figure A1.2- Plot of USGS Surface Water Daily Data, and a daily aggregation of USGS Surface Water Historical Observations (15-minute data).

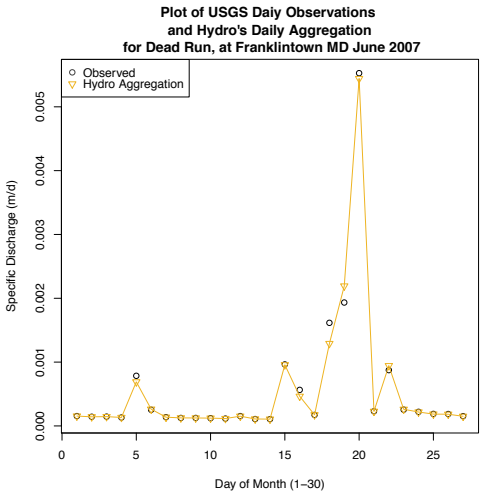


Figure A1.3- Plot of USGS Surface Water Daily Data, and a daily aggregation of USGS Surface Water Historical Observations performed by Hydro (1st observation of every hour)

All streamflow data for this report were obtained from the USGS. Five sites located in and around Baltimore, MD were considered for this study: Dead Run at Franklintown, MD (USGS Gauge #01589330), Baisman Run at Broadmoor, MD (USGS Gauge #01583580), West Branch Herring Run at Idlewylde, MD (USGS Gauge #01585200), East Branch Herbert Run at Arbutus, MD (USGS gauge #01589100), and Gwynns Falls near Delight, MD (USGS gauge #01589197). These sites were chosen because: (1) they are all located relatively close to each other, (2) they have a range of impervious cover, (3) Dead Run was previously used as a case study for Hydro (Wang et al. 2008), and (4) the availability of unbiased NEXRAD precipitation data for these sites (Smith et al. 2013). Watershed characteristics of these sites are listed below in Table A1.1, including the USGS Gauge number, drainage area (mi^2) and land cover percentages for the 5 study watersheds. Figure A1.4 provides a map of the Baltimore city limits, and watershed boundaries for the gauging stations at the study sites. Watershed boundaries for the study sites were obtained using the USGS program StreamStats (USGS 2016).

Land cover classifications for the five watersheds were completed by Charity Nyelele, a PhD student at SUNY ESF who is part of our i-Tree team. Tree, short vegetation, bare soil, water and impervious cover percentages were obtained from a 3.2ft Urban Tree Canopy (UTC) land cover dataset from the USDA Forest Service's UTC assessment (O'Neil-Dunne, 2018). Runs of Hydro using the GUI were performed with a directly connected impervious area (DCIA) that was calculated using the Sutherland equation within i-Tree Hydro framework (Sutherland, 2000). For the fraction of trees over pervious area, fraction of trees over impervious area, fraction of short vegetation in the catchment, fraction of bare soil, fraction of evergreen trees and fraction of evergreen shrub (short vegetation), the values based on image classification and the National Land Cover Database (NLDC) 2011 data were used and adjusted proportionally based on the land cover percentages in each watershed (USGS 2011).

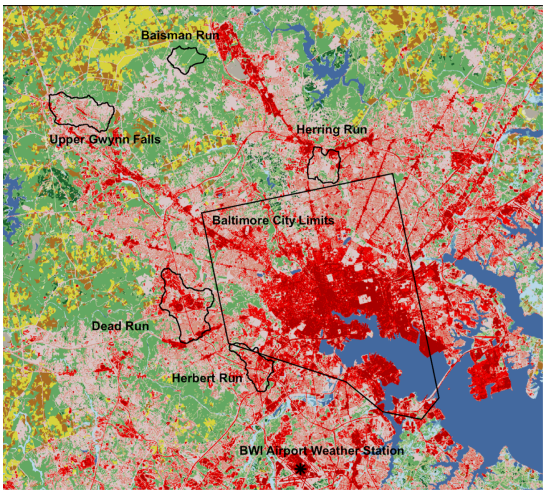


Figure A1.4- Study Site's watershed boundaries and NLCD 2011 land cover.

3.0 Morris One-At-a-Time (MOAT) Sensitivity Analysis

Sensitivity analyses (SA) are general techniques to assess how input variables (independent variables) impact model output (dependent variables). There are many techniques that can be employed depending on the model structure, the interaction between variables, and the model output.

Table A1.1- USGS Gauge Number, drainage area, and land cover percentages for the five Study Sites

Land Cover Type	Baisman Run	Upper Gwynn's Falls	Herring Run	Herbert Run	Dead Run
USGS Gauge Number	01583580	01589197	01585200	01589100	01589330
Area (mi^2)	1.47	4.23	2.13	2.47	5.52
Trees	79.9%	41.8%	43.7%	32.8%	27.7%
Short Vegetation	16.3%	30.0%	23.7%	21.3%	19.7%
Impervious Cover	3.7%	28.1%	32.1%	45.1%	52.2%
Water	0.0%	0.0%	0.2%	0.4%	0.2%
Bare Soil	0.0%	0.1%	0.2%	0.4%	0.1%
Trees Over Impervious	7.3%	3.8%	4.0%	3.0%	2.5%
Trees Over Pervious	72.5%	38.0%	39.7%	29.8%	25.2%
Evergreen Trees	0.0%	1.5%	2.4%	4.3%	0.0%
Evergreen Shrub	0.5%	0.9%	0.0%	0.0%	0.1%
DCIA	2.36%	18.09%	21.79%	42.16%	48.86%

Here we wish to assess which input parameters Hydro is most sensitive to, in an effort to aide model builders and users. For this assessment, Hydro was run for three years, from January 1, 2007 to December 30, 2009 for Dead Run at Franklintown, MD. This site was chosen because of past Hydro development here (Wang et al. 2008). With the computing power available, Hydro took roughly 46 seconds to complete this 3-year simulation. As such, selecting an algorithm that was less computationally intensive was a primary objective. The main goal for the SA was to identify a subset of the parameters that may be employed to calibrate Hydro in the future.

Another area of concern is the large ranges given to several model parameters (T0 is one such example with a lower limit is 0.0005 ($\frac{m^2}{h}$), and an upper limit of 150 ($\frac{m^2}{h}$)). While this does make the model applicable to watersheds with vastly different characteristics, it is believed that sometimes parameters

may get "stuck" in an area of the solution space that is far from an "optimal" value, which overall, may have a negative impact on the calibration. A secondary goal of the SA, therefore, was to aid in the identification of parameters whose ranges may need to be refined. Additionally, we believe users should potentially be guided towards employing more realistic parameter ranges based on the characteristics of their given study area. If the user does not have the expertise to create these ranges, this could be done via different default settings a user could select or based on watershed location.

With the above defined goals, a Morris One-At-a-Time (MOAT) algorithm was selected for this sensitivity analysis (Morris, 1991). MOAT is a global SA tool that works to determine which parameters have a large effect on model output across a wide range of conditions. One drawback is that this method assumes independence among model parameters, which is not always the case in highly parameterized hydrologic models. The algorithm starts by randomly sampling k parameters at predefined levels (x_1, . . ., x_k) and determining the output of the model (y). One of these parameters, x_i, is then increased by a step size, Δ, while the other parameters are held at starting values, and the model is then rerun with this new parameter set. From this, one can estimate the elementary effect for parameter x_i, ee_i:

$$ee_i(X) = \frac{\frac{y(x_1,\ldots,x_{i-1},x_i+\Delta,\ldots,x_k)-y(X)}{y(X)}}{\frac{\Delta}{x_i}} \quad \text{(A1.2)}$$

where y(X) is output from a model run with $X = (x_1, x_2, \ldots, x_k)$. The elementary effect represents the percent change in the output due to a percent change in the input parameter. For this assessment, 27 parameters were considered. The range of each parameter was divided into 5 intervals from which sampling occurred. 100 sets of elementary effects were determined for each parameter, randomly sampling the starting values for each iteration. Descriptions of the parameters included in the SA are presented in Table A1.2. Two different output scenarios (for y) were considered in the sensitivity analysis. The first output scenario was the total volume of runoff (total flow), which was used to identify parameters that impact evapotranspiration. The second output scenario was the total volume of surface

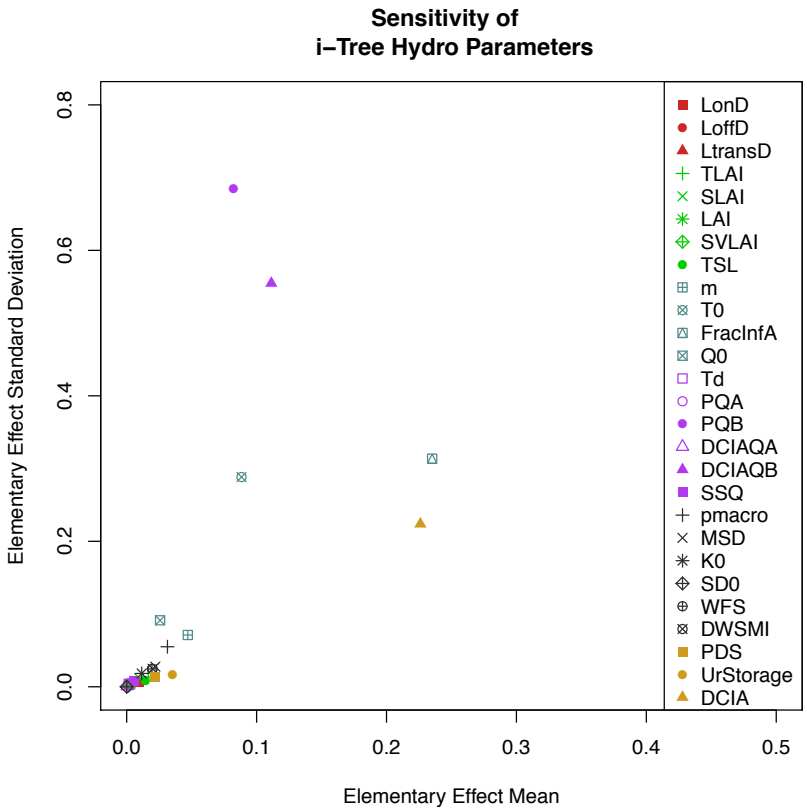
runoff (surface runoff). The surface runoff scenario was used to determine the most important parameters that impact surface runoff processes.

Using the above analysis, 100 values of the ee were generated for each parameter, and the sample mean and standard deviation of these 100 values were determined. A parameter with a high mean ee has a larger effect on the model output than a parameter with a low mean ee. A high ee standard deviation indicates that the effect of the respective parameter changes across the parameter space, or that there are interaction effects between the respective parameter and other parameters in the model. A low ee standard deviation indicates that the impact the respective parameter has on model output is relatively constant across the parameter space. Figures A1.5a and A1.6a present the output from the SA performed on the total flow and surface runoff scenarios, respectively. For each scenario, there appears to be a relatively small subset of parameters that have a disproportionately high effect on model output. There also appears to be a tightly clustered subset of parameters at the lower end of the ee range (for both the mean and standard deviation) for both output scenarios. Figures A1.5b and A1.6b present a reduced range of SA output so the reader can more easily distinguish between variables on the lower end of the range of the mean and standard deviation of ee.

The five parameters with the highest mean ee for the total flow scenario are fraction of the catchment generating infiltration excess overland flow (FracInfA), directly connected impervious area (DCIA), the beta time constants for both pervious and directly connected impervious flow (PQB and DCIAQB, respectively), and the transmissivity at saturation (T0). The five parameters with the highest mean ee for the surface runoff scenario are DCIA, FracInfA, T0, DCIAQB (which were also identified for the total flow scenario), and the soil macropore percentage (pmacro). It should be noted that the impact of pmacro was considerably less than the impact of the 4 aforementioned parameters in the surface runoff scenario.

Table A1.2- Hydro variable names and descriptions, and our own personal naming convention

Description in param.dat	Variable in param.dat	Variable in SA Legend	Parameter Range
Leaf on date of the year	LeafOnDay	LonD	1-365
Leaf off date of the year	LeafOffDay	LoffD	1-365
Length of leaf transition days	LeafTransDays	LtransD	1-365
Average LAI of tree bark in the catchment	TreeBarkLAI	TLAI	1-- 7
Average LAI of shrub bark (short vegetation) in the catchment)	ShrubBarkLAI	SLAI	1--7
Specific leaf storage	TreeSL	TSL	No range
Scale Parameter Soil Transmissiviy	m	m	0.01 - 1.2
Transmissivity at Saturation	T0	T0	0.0005-150
Unsaturated zone time delay	Td	Td	0.01-150
Time constant for pervious area flow	PAQ_RT_A	PQA	0.01-240
Time constant for pervious area flow	PAQ_RT_B	PQB	0.01-240
Time constant for directly connected imperveous flow	DCIAQ_RT_A	DCIAQA	0.01-240
Time constant for directly connected imperveous flow	DCIAQ_RT_B	DCIAQB	0.01-240
Time constant for subsurface flow	SSQ_RT	SSQ	0.01-240
Soil Macropore Percentage	pmacro	pmacro	0.0000001-0.2
Fraction of catchment generating infiltration excess overland flow	FracInfA	FracInfA	0-1
Initial root zone deficit	SD0	SD0	0-1
Maximum root zone storage deficit	MSD	MSD	0.001-1
Wetted moisture content	DWSMI	DWSMI	0.1-0.7
Wetting front suction	WFS	WFS	0.03-0.4
Initial stream discharge	Q0	Q0	0.00000001-0.05
Average catchment tree leaf area index in summer	TreeLAI	LAI	1--7
Average short vegetation LAI	ShortVegLAI	SVLAI	1--7
Imperveous Depression Storage	UrStorage	UrStorage	0-1
Pervious Depression Storage	PDS	PDS	0.1-7.5
Fraction of connected imperveous area (IA) of total IA	FracConImp_Timp	CIA	0-1
Surface hydraulic conductivity	K0	K0	0.01-100



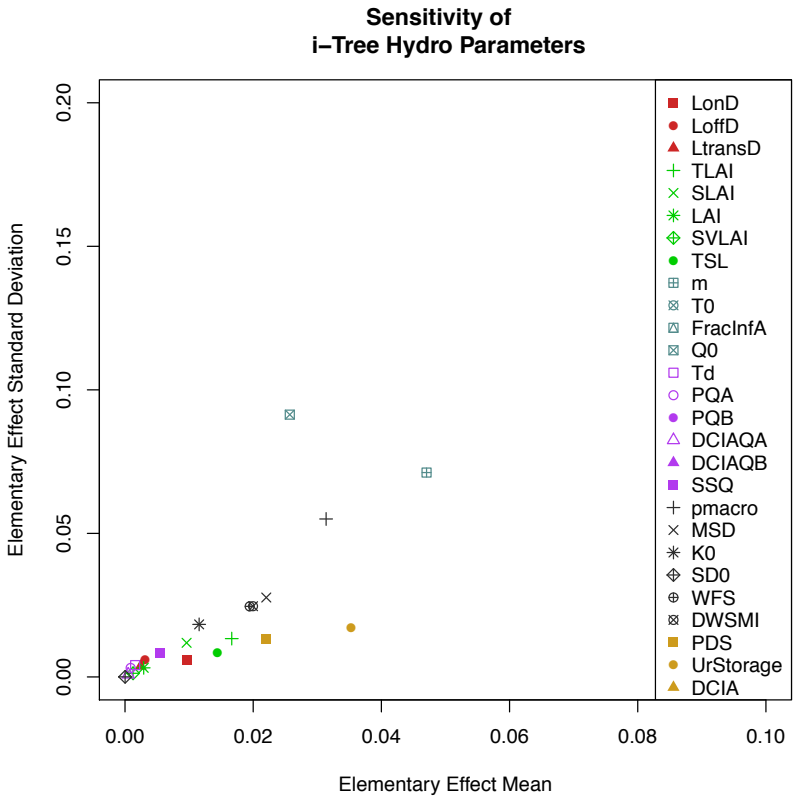


Figure A1.5- Full (top) and constrained (bottom) range SA output for total flows

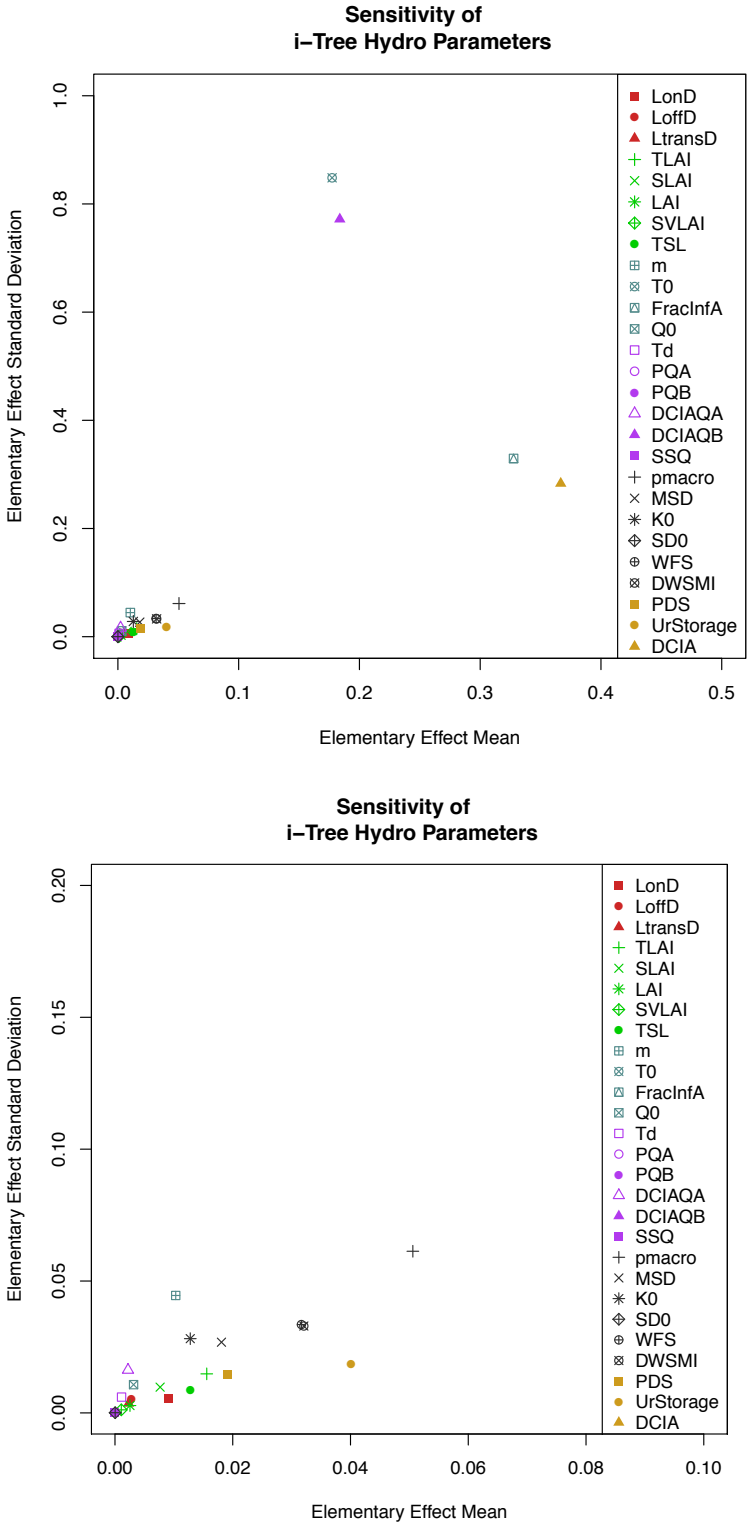


Figure A1.6- Full (top) and constrained (bottom) range SA output for surface runoff

The output from the SA was insightful for several reasons. The PEST routine for Hydro currently calibrates on the maximum root zone storage deficit (MSD), the scale parameter for soil transmissivity (m), pmacro, and T0. While T0 does appear to have a larger effect on model output (based on our SA results), the other three parameters appear to have a minimal impact on model output. For the total flow scenario, MSD, pmacro and m had a mean ee on the lower end of the range. Note that pmacro and m had a mean ee that is considerably larger than that of MSD; however, the effect that all three of these parameters have on model output appears to be negligible. Similarly, these three parameters were also determined to have relatively small mean ee for the surface runoff scenario. MSD and m were generally among the parameters with the lowest mean ee, while pmacro had a slightly larger mean ee.

One surprising outcome of the SA is the relative insignificance of parameters related to leaf area index (LAI). These parameters were determined to have the smallest impact on model output for both scenarios considered. For end users, acquiring accurate at-site estimates of LAI's seems to be far less important that understanding the impervious area and the soil characteristics of the site. This is a surprising result, as one would expect LAI to have an impact of ET, and thus be a more important parameter for the total flow SA.

Further, it is believed that correctly quantifying the initial stream discharge (Q0) is very important for model fitting, especially for shorter simulations. Currently Q0 is determined from either observed streamflow or as a default value that is a linear function of drainage area, a model form that is not consistent with current log-linear models recommended in the literature for predicting mean annual streamflows and streamflow quantiles (Vogel et al. 1999). We encourage the i-Tree Hydro development team to update the model such that this parameter is determined in a manner more consistent with typical techniques employed to predict streamflow quantiles in the US.

One aspect that was not explored through the SA was the importance of precipitation input data. The SA was conducted with bias corrected Next Generation Radar (NEXRAD) precipitation data provided by Dr.

Brianne Smith of CUNY-Brooklyn. The Hydro GUI provides automated links to many different weather stations. The closest weather station to the study sites is Baltimore Washington International Airport (BWI). From previous work by Jermyn (2018), it was determined that the NEXRAD data provides much better model fitting than does the BWI data. This was confirmed through our work in assessing the Hydro calibration routine (see Table A1.3 in Section A1.4.0 of this document). For these reasons, NEXRAD precipitation was employed in the SA.

In summary, it is believed that moving towards a calibration scenario that incorporates the parameters FracInfA, DCIA and the beta time constants will improve calibrated model fitting. Additionally, understanding the impact that precipitation data has on the simulation scenarios appears desirable. Both of these scenarios are explored in the next section of this document.

4.0 Alternative Calibration Scenarios

This assessment now moves towards investigating the Parameter Estimation and Uncertainty Analysis package (PEST), which is currently used to calibrate Hydro. The first step in this analysis was to perform PEST calibrations using the Hydro GUI at all 5 study sites: Dead Run, Baisman Run, Herring Run, Herbert Run, and Gwynns Falls. This served as a baseline for further analyses and was completed using both the NEXRAD and BWI precipitation data. For all simulations, Hydro was calibrated on weekly flows over the year 2008 (from January 1 to December 30). Use of weekly flows as opposed to a shorter time step was employed because here we are most concerned with the model's ability to capture the water balance, as opposed to fitting peak flows. For each site, the NSE is reported for a run of Hydro with the suggested default values of the four parameters Hydro currently calibrates on (T0, m, MSD, and pmacro); this simulation is referred to as Default in Table A1.3 and is the uncalibrated model results. Table A1.3 also presents the weekly NSE from a run of Hydro after those parameters have been calibrated using PEST (referred to as Calibration).

Table A1.3- Weekly NSE from Hydro Suggested Default Parameter Values and PEST Autocalibrated Parameter Values for the Five Study Watershed's

Site	Dead	Baisman	Herring	Herbert	Gwynns
Default BWI	0.120	-55.4	-0.911	-1.9	-6.9
Calibration BWI	0.580	-4.4	0.715	0.586	-0.156
Default NEXRAD	0.625	-67.3	-1.2	-0.749	-4.3
Calibration NEXRAD	0.817	-8.7	-0.271	0.701	0.302

By far the best model fitting occurred at Dead Run. It is interesting that fitting at Dead Run was substantially better than fitting at Herbert Run. Both watersheds have relatively similar characteristics, so the expectation would be that model fits at each site would be similar. This is not the case, and it appears that the suggested default values are better suited to provide accurate fitting at Dead Run. Regardless, an encouraging result is that the PEST autocalibration routine unanimously produced better fits than the uncalibrated Hydro with the suggested default parameter values. This finding was consistent at all sites, for both the NEXRAD and BWI precipitation datasets.

The NEXRAD precipitation data generally resulted in better fits when compared to fitting with BWI precipitation data. Two exceptions are at the sites Baisman and Herring Run. The fits at Baisman Run were particularly poor irrespective of the input precipitation data. It is interesting that Herring Run fits better to the BWI data, considering it is geographically one of the more remote sites from the BWI airport (see Figure A1.4). One would expect the BWI and NEXRAD data to produce the most similar fitting at Herbert Run, due to its close proximity to the airport, although this does not appear to be the case. The expectation that the bias corrected NEXRAD precipitation data provides better model fitting is generally met (particularly for the default (i.e. no calibration) runs). Therefore, in some situations it may be worth the extra effort to obtain more site-specific highly localized precipitation data as an input for running

Hydro. The results in Table A1.3 clearly indicate the benefits of model calibrate over the use of the default parameters in terms of model fit.

The remainder of the paper will focus on two sites (Dead Run and Herbert Run) and one precipitation dataset (NEXRAD precipitation data). Hydro performance was "best" at these two sites with NEXRAD precipitation based on the NSE of weekly streamflows. Additionally, Dead Run was selected as this was the watershed used by Wang et al. (2008) in their Hydro case study. Herbert Run was selected for two additional reasons. First, the calibration routine seemed to be of considerable benefit at this site. The runs with the suggested default parameter values for both precipitation datasets produced negative weekly NSE's (-1.9 and -0.749 for the BWI and NEXRAD precipitation datasets, respectively), which increased to positive NSE's (0.586 and 0.701 for BWI and NEXRAD precipitation datasets respectively) after calibration. Herbert was also selected on the basis that model fitting poorer that fitting at Dead Run despite the similarity in site characteristics between the two watersheds. Exploring these two sites could highlight weaknesses of the model, problems with the suggested default parameter values, or potential inadequacies of the current Hydro calibration routine.

For the remainder of this analysis, calibrations are performed in R. These calibrations used USGS daily average streamflow as observations, which were aggregated to the desired timestep (i.e., the Hydro processed streamflow is no longer being used as the observations). The Hydro solution space appears to be a difficult solution space to work with, resulting in unforeseen challenges in model calibration. We considered four fundamentally different optimization approaches, three of which did not provide reasonable results. A Gauss Newton Levenberg Marquardt (GNLM) algorithm was employed first. The motivation behind selecting a GNLM was its similarity to the routine currently implemented in PEST. Our routine uses the R package minpack.lm. After much initial testing, it was determined that minpack.lm was not appropriate for this study as we had trouble producing similar calibrations to those from PEST within Hydro model. A Genetic Algorithm (GA) was then considered, as such a routine represents a

fundamentally different approach to non-linear optimization. The GA used was a machine coded genetic algorithm (MCGA) obtained from the R package mcga (Satman 2018). The computational requirement for this package was heavy and yielded inconsistent results (final calibrated results that were sometimes better than PEST, and sometimes worse than PEST). A simulated annealing (SA) approach was considered next. Such an algorithm could potentially be implemented with Hydro more easily than the GNLM and GA algorithms due to its typically increased efficiency and robustness. The R package GenSA (Xiang et al. 2013) was used; however, like the GA, the results from this approach were determined to be inconsistent. Note that Appendix A1.B contains a more detailed discussion of these calibration routines.

As a final effort, the shuffled complex evolutionary (SCE) algorithm, proposed by Duan et al. (1992) for hydrologic model calibration, was employed to calibrate Hydro. Dettmann et al. (2018) developed the R package SoilHyP, containing the function SCEoptim which is an R adaptation of the original Matlab code used by Duan et al. (1992). This algorithm combines the strengths of the simplex procedure with the concepts of a controlled random search, aided by the competitive complex evolution (CCE) algorithm (Duan et al. 1992). Due to its extremely high computational demand (i.e. lots of model runs are required), the SCE algorithm was not initially considered for calibration of Hydro, but given the shortcomings of the three aforementioned algorithms, implementation of SCE was explored. Because of the increased computational requirements, the simulation period was modified. The three-year simulations (used for the sensitivity analysis) were no longer feasible due to computational time limitations, and instead a simulation period of a one year was used. For all ensuing calibrations (PEST included) the simulation period was from 1 January 2008 to 30 December 2008 (52 weeks). The number of iterations for the SCE was set to a maximum number of 150, although only about 45 were used for each calibration before achieving convergence.

The major advantage of performing the calibrations in R was the flexibility to implement variations from the standard PEST routine. An alternative objective function and two different parameter subsets were investigated in this calibration experiment. An alternative objective function, one that minimizes the sum of squares of the monthly runoff ratio, is explored. Calibrating on this metric may move parameter values towards a region of the solution space so that the resulting model better represents the water balance, and thus the impact of trees on urban hydrologic processes. We define the monthly runoff ratio as the ratio of monthly streamflow to monthly precipitation. The monthly runoff ratio provides a general assessment of how well the model partitions precipitation as runoff and evapotranspiration, assuming that the differences in changes in storage between the model and the watershed are negligible compared to the precipitation, runoff and ET fluxes over the monthly time step (which may not always be valid). The two parameter subsets considered for calibration will be referred to as PEST4 and SEN4. PEST4 calibrates the four parameters on which Hydro currently calibrates (T0, pmacro, MSD, and m). All other parameter values are set to Hydro default values. SEN4 calibrates the four parameters identified from the sensitivity analysis as being most important to the total runoff (T0, DCIA, DCIAQB and FracInfA). Again, all other parameters are set to Hydro default values.

Output from these calibrations can be seen in Table A1.4. This table presents the weekly NSE (wNSE) and monthly runoff ratio NSE (rrNSE) for several runs of Hydro. The first month of the simulation (January 2008) was used as a "ramp-up" period; the statistics in Table A1.4 are for the 11 remaining months of the year (February 2008 to December 2008). Results are presented for the default case (i.e. no calibration) and various calibrations that maximize the NSE of weekly flows and monthly RR's at both Dead and Herbert Run. Regardless of the objective function, or optimization algorithm used, the wNSE and rrNSE were calculated, to allow for assessment of the tradeoffs between the two objective functions being considered. Note that Appendix A1.C contains the observed and predicted weekly streamflow and month runoff ratios for all of the simulations in Table A1.4.

First the wNSEs and rrNSEs are presented for the default case and a Hydro GUI calibration using PEST. These serve as baseline to compare the alternative calibration scenarios against. The PEST4 parameters were then calibrated employing SCE, considering both objective functions. These calibration scenarios were also performed for the SEN4 parameters.

Table A1.4 is insightful for many reasons. In general, model fits (as measured by both NSE metrics) were better at Dead Run than model fits at Herbert run, a finding that was previously observed. At both sites, when calibrating on the weekly flows, the SCE always found a higher wNSE than was found with PEST, indicating that the SCE algorithm is not only effectively searching the solution space, but also doing a better job than the current version of PEST in Hydro. Additionally, a switch from the PEST4 parameters to the SEN4 parameters resulted in further increases to the wNSE at both sites, indicating that a change in calibration parameters may improve model performance. This result also provides further support to the results of the sensitivity analysis (SA).

Table A1.4- Weekly Flow NSE and Monthly RR NSE from the Alternative Calibration Scenarios

		Dead Run		Herbert Run	
Scenario (with OF)	Calibration Routine	Weekly Flow NSE	Monthly RR NSE	Weekly Flow NSE	Monthly RR NSE
Default Values (weekly flows)	n/a	0.657	-1.81	-0.71	-11.81
PEST 4 (weekly flows)	PEST	0.835	0.122	0.718	-1.27
	SCE	0.865	0.332	0.863	-0.494
PEST4 (monthly RR)	SCE	0.86	0.38	0.826	-0.474
SEN4 (weekly flows)	SCE	0.943	0.509	0.874	-0.779
SEN4 (monthly RR)	SCE	0.935	0.546	0.707	-0.522

Of most concern in this analysis, though, is the poor fits of Hydro to the monthly runoff ratios. While we realize that the short simulation period and the lack of a longer "ramp-up" period prior to the calibration may impact these results, it appears that Hydro does a very poor job representing runoff ratios and thus may not capture the processes important to urban hydrology. At Herbert Run, the rrNSE is always less than 0, indicating that the model provides worse estimates of the runoff ratio than the arithmetic mean of the observed runoff ratios. Note that this happens even when the calibration is trying to maximize the rrNSE. At Dead Run, the rrNSE is only greater than 0 when considering the SEN4 parameters.

The implications of each of these findings for the i-Tree Hydro development team are different. Currently, end users are encouraged not to calibrate, due to the computational requirements to implement PEST. As shown previously (see Table A1.3), without calibration Hydro often produces extremely poor model fits to the observations. It is believed that PEST previously calibrated Hydro using an algorithm similar to the SCE. While this algorithm has considerably higher computational requirements than does the current calibration in PEST (which is based on the GNLM algorithm), it appears the SCE algorithm does a much better job identifying more optimal parameter sets. For our study, we calibrated over a single year, which on our computers typically required 16-23 hours for the SCE calibration routine to complete. A single year is a short simulation period, and the computational demands required to implement the SCE algorithm over a longer simulation period might outweigh the benefits, to the end user, of the improved fitting the SCE provides.

Additionally, our SCE algorithm occasionally required multiple runs to "find" optimal solutions. At both Dead and Herbert Run, runs were initially performed with the suggested Hydro default parameter values. When considering the SEN4 parameter subset, calibrations on the monthly RR sometimes produced higher wNSE's and rrNSE's than did calibrating on the weekly flows. This was an indication that the SCE algorithm was also getting "stuck" and finding a local, and not a global optimum, when calibrating on the weekly flow objective function with the SEN4 parameters. As such, the final parameters from the

monthly runoff ratio scenario were used to initialize a second run with the weekly flows objective function. This was performed at both sites, and allowed the SCE to find better optimal parameter values. This again points to issues with the shape of the parameter solution space in Hydro, which may make calibration of this model difficult.

A more feasible approach would be to change the subset of parameters on which Hydro currently calibrates. Unfortunately, we could not run the SEN4 parameters through the current PEST algorithm. However, when considering the weekly flow objective function, a switch to calibrating on the SEN4 parameters from the PEST4 parameters resulted in higher wNSE. If calibration is moved towards some combination of those parameters identified above in Figures A1.5 and A1.6, model fitting should improve. We believe the i-Tree Hydro development team should actively work towards updating the parameters on which the model calibrates to reflect the results of this SA. At the very least, the parameter DCIA should be incorporated into the current calibration routine, as this was one of the most sensitive model parameters for both SA scenario's (total flow and surface runoff).

Another important finding is the differences in the rrNSE between the two sites. At Herbert Run, the PEST4 parameters provide better fits to the monthly runoff ratios than do the SEN4 parameters, though both are less than 0. This finding is the opposite at Dead Run, where the SEN4 parameters provide better fits to the monthly runoff ratios than do the PEST4 parameters. This may be an indication that the suggested default parameter values are negatively impacting calibration at Herbert Run. This is intriguing due to the similarity of both sites, as one would expect similar fitting at both Dead and Herbert Run with similar parameter values. Since a previous study has used Dead Run (Wang et al. 2008), perhaps the default parameters were chosen to fit this site. Confirming this observation will require extensive modeling review and simulation and is beyond the scope of this analysis. One idea to potentially alleviate this issue would be to move towards constrained parameter ranges based on specific at site conditions. The parameters exerting control on soil processes seem to be the most sensitive, and Hydro provides users

with the ability to select specific soil types each having their own set values for surface hydrologic conductivity (K0), wetting front suction (WFS), and wetted moisture content (DWSMI). One suggestion might to be to reduce the ranges for parameters like T0 or the time constants (e.g., PAQ_RT_A, PAQ_RT_B, DCIAQ_RT_A, DCIAQ_RT_B), whose parameter ranges span multiple orders of magnitude, to be more consistent with what their values might be for given soil types.

It is also entirely possible that the difference in fitting between the two sites is less an issue with potentially poor default parameter values, and more a consequence of the model fitting poorly in very small watersheds. These two sites do differ considerably in size. Dead Run has a drainage area of 5.52 mi^2, which is more than twice the drainage area of Herbert Run (2.47 mi^2). This may be a cause for some concern as many urban watersheds are small in size.

Regardless, the poor fits to the monthly RR is of concern. The rrNSE's were consistently below 0 (except for at Dead Run with the SEN4 calibration parameters). On the other hand, the NSE's of the weekly flows were remarkably high (up to a wNSE of 0.94 at Dead Run for the SEN4 parameters), indicating very good agreement between modeled and observed streamflow. This was observed for both the SCE and PEST calibrations. This finding was surprising, although important. Considering these poor fits, we believe there may be an issue with the model, particularly surrounding its ability to accurately capture ET (which is the primary goal of this model). We believe this is a critically important issue for the i-Tree Hydro development team to further investigate these findings.

5.0 References

Dettmann, U., Andrews, F., Donckels, B., & Duan, Q. (2018). Package 'SoilHyP'.

Elzhov, T. V., Mullen, K. M., Spiess, A. N., Bolker, B., Mullen, M. K. M., & Suggests, M. A. S. S. (2016). R Package 'minpack. lm'.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161-174.

O'Neil-Dunne, J. (2018). Tree Canopy Assessment. Retrieved April 1, 2018 from http://gis.w3.uvm.edu/utc/.

Satman, M. H. mcga: Machine coded genetic algorithms for real-valued optimization problems (2018). *R package version 3.0*, *3*.

Smith, B. K., Smith, J. A., Baeck, M. L., Villarini, G., & Wright, D. B. (2013). Spectrum of storm event hydrologic response in urban watersheds. Water Resources Research, 49(5), 2649-2663. doi:10.1002/wrcr.20223.

Sutherland, R. C. (2000). Methods of Estimating Effective Impervious Area of Urban watersheds. Watershed Protection Techniques,2(1), 282-284. Retrieved April 12, 2018.

United States Geological Survey: Scott Phillips, personal communication, June 2018.

U.S. Geological Survey, 2016, The StreamStats program, online at http://streamstats.usgs.gov, accessed between (January to April 2018).

U.S. Geological Survey, 20141010, NLCD 2011 Land Cover (2011 Edition, amended 2014) - National Geospatial Data Asset (NGDA) Land Use Land Cover: U.S. Geological Survey.

Vogel, R.M., I. Wilson and C. Daly, Regional Regression Models of Annual Streamflow for the United States, *Journal of Irrigation and Drainage Engineering*, ASCE, 125(3), 148-157, May/June, 1999.

Wang, J., Endreny, T. A., & Nowak, D. J. (2008). Mechanistic simulation of tree effects in an urban water balance model 1. *JAWRA Journal of the American Water Resources Association*, *44*(1), 75-85.

Appendix A1.A – List of Suggested Improvements to i-Tree Hydro (based on Hydro version 6.1.1)

1) Improvements to the GUI
    a. Project Area Information
        i. Change “Browse for my own topographic data” under Topographic Data to “Select my own topographic data” as this is more consistent with the verbiage used in Weather Station Data, and Calibration Data sections of the GUI.

    b. Land Cover Inputs
        i. No issues.

    c. Hydrological Parameter’s
        i. Shrub Bark and Tree Bark LAI—this is confusing as all other LAI parameters are in Step 2- Land Cover Inputs.
        ii. Pervious and Impervious Depression Storage, and Surface Hydraulic Conductivity seem to be handled in the section of the code that deals with the Land Cover parameters (LandCover.xml). This was also somewhat confusing.

    d. Run Hydro Model
        i. No issues.

2) Parameters
    a. K0—This is listed as Soil Hydraulic Conductivity and has a range of 0.01-100. The GUI turns this into a fraction of a percent meaning the range acceptable by the code is 0.000001- 0.01. After a brief review of the code (in July of 2018), Robert Coville (of i-Tree Hydro development team) and Brenden Covert (coauthor of this report) were unable to determine how specifically this parameter is handled.

    b. Pmacro— It is known there is a potential bug with this parameter (personal communication with Robert Coville).

    c. MSD— We believe this parameter is somewhat misidentified in the GUI. MSD in this version it is described to the user as "Depth of Upper Soil Zone", but based on how we believe the parameter to be used in the code, it is more specifically the “depth of water column (available water space in 1-D) in the upper soil layer which is active in evapotranspiration." Somehow this should be clarified in the user interface and help text, or if this is intended to be a soil depth, then MSD should also account for soil's drainable porosity.

    d. We would recommend a review of all other parameters to ensure their accuracy, with particular emphasis on parameters whose ranges may need to be refined (i.e. T0, any of the time constants etc.).

    e. TSL has no defined parameter range in this version. This parameter should be updated to have a defined range.

3) General Issues
    a. Hydro would benefit considerably from improved documentation of the model. While some papers in the literature provide information regarding routines within Hydro, it is our understanding that there is no single document that outlines the model structure and

methods. Such a document would greatly help model users improve their understanding and use of model output.

b. The extended outputs of the model take a very long time to write. This increases the amount of time to complete a run of Hydro. Currently the only way to turn these off is to manually modify the "hydro_config_file.XML." We would recommend, if possible, to add this functionality to the GUI.

c. In the "hydro_config_file.XML" there is a line of code establishing run types 1-4, and route types 1-3. It was noticed by happenstance by coauthor Brenden Covert that changing these inputs in the hydro_config_file.XML significantly impact model output. Coauthor Brenden Covert has tried but could not find any literature on what these run and route types are. Users should be provided some explanation of what these are, as changing them may benefit their simulation.

4) Additional Suggestions for Model Improvement
   a. Use daily USGS streamflow as a model input for Hydro calibrations instead of the current methodology of aggregating daily streamflow in the model.

   b. Allow users the ability to set or reduce the allowable range of parameter values based on site characteristics.

   c. Change the parameters on which the model is calibrated. The SEN4 parameters identified in this document are a good starting point, but this should be further investigated.

   d. Update the methodology to estimate initial stream discharge to be more consistent with accepted methods in practice.

   e. Perform additional model testing and verification at a variety of watersheds to assess model performance and its ability to represent the water balance and the impact of changing tree and impervious cover.

   f. Further explore the impact and performance of model calibration, considering alternative objective functions and calibration methods.

Appendix A1.B – Issues with Calibration Routines

The discussion below is provided to the i-Tree Hydro development team to help them further investigate model calibration. Considerable efforts were made to calibrate Hydro. Should Hydro calibration be further explored (which we recommend).

The Gauss Newton Levenberg Marquardt algorithm was implemented first as this is a very similar optimization approach to PEST. Ideally, results from this routine should closely resemble the results from PEST if either algorithm is adequately robust for use in the Hydro parameter solution space. The R package minpack.lm has the function nls.lm which is designed for solving nonlinear least-squares problems employing a modification of the Levenberg-Marquardt algorithm (Elzhov et al. 2016). For this routine, a three-year simulation period was used (January 1, 2007 to December 30, 2009). Note that this was the same simulation period used in the sensitivity analysis. All routines were initialized with the Hydro suggested default parameter values. It was found that this routine often converged and terminated in roughly 20-35 minutes. For reference, a typical PEST calibration over the same simulation period typically achieved convergence in roughly 40-60 minutes. While this routine was fast, and presumably efficient, it was deemed unfit for use with the Hydro solution space. The final autocalibrated parameter values consistently moved towards extreme ends of the parameter space. This routine never provided reliable results, and consistently produced poor NSE's.

A machine coded genetic algorithm (GA) was the most computationally efficient of the several GA algorithms explored. The R Package mcga uses a byte representation of variables, which is the driver of that increased efficiency. The GA used had built in control for parameter bounds, which allowed for the calibration to remain within the same bounds that PEST imposes (i.e. the Hydro defined parameter bounds). This represented an improvement on nls.lm where functionality to handle parameter bounds was implemented manually (i.e. through the code developed in R that was used to run Hydro). We first explored changing the number of iterations and a constrained simulation period. Given the known

increased computational demands, it was deemed infeasible to run the GA over the full 3-year simulation period that was used for the initial SA. Initially, a 3-month run simulation was considered with the number of iterations set to 350. This run took roughly 5 days (120 hours) to complete. The next step was to increase the simulation length and decrease the number of iterations to see if we could arrive at some middle ground. After a series of initial tests, a one-year simulation period, with 15 iterations was used. It had been determined that 15 iterations are sufficiently adequate to explore the solution space (larger numbers of iterations generally produced similar results). Considering the computational requirements to implement this procedure and the relatively long run time of Hydro (roughly 17 seconds for a one-year simulation), selecting the smallest number of iterations that provided sufficient results was important. The GA was run with an initial population of 100, a relatively high mutation rate of 0.25 (to aid in the full exploration of the solution space), and elitism (which always keeps the best simulations from the population in the next iteration). A typical run time for our simulation period and initial conditions was anywhere from 12-18 hours. This algorithm, though, had trouble identifying optimal parameters for the rrNSE. It was consistently seen that calibrations considering an objective function of the weekly flows produced higher rrNSE's than calibrations on the monthly RR objective function. This proved to us that the algorithm was getting stuck in poor regions of the solution space and could not be used to effectively calibrate Hydro.

Given the extensive run times, rather than trying to tailor the GA for use with Hydro, it was determined that a more efficient calibration algorithm should be explored. Simulated annealing begins in some region of the solution space (for our use, the suggested default parameter values), which is the initial state. Then with some probability, the system moves to some other state. This probability is typically modeled with a form of a Cauchy distribution (our algorithm employed the distorted Cauchy-Lorenz distribution). The probability of moving away from the current state is controlled by the temperature where a higher temperature results in a higher probability that the system changes state and is called the annealing schedule. As the annealing schedule proceeds, the temperature decreases. Ideally, at the start of the

algorithm, one would prefer to make "bad" jumps (move to a worse portion of the solution space), associated with high temperatures such that the solution space is adequately explored. As the temperature lowers, through the procession of the annealing schedule, the hope is that one arrives closer to a globally optimal solution. This process typically finds optimal values much faster than the GA. These runs typically took 8-10 hours. Unfortunately, we experienced similar struggles with simulated annealing as we had experienced with the GA, often not converging to optimal values. This moved us to our final optimization routine, the shuffled complex evolution algorithm, which was used in this analysis.

Appendix A1.C – Time Series Plots of Hydro Output

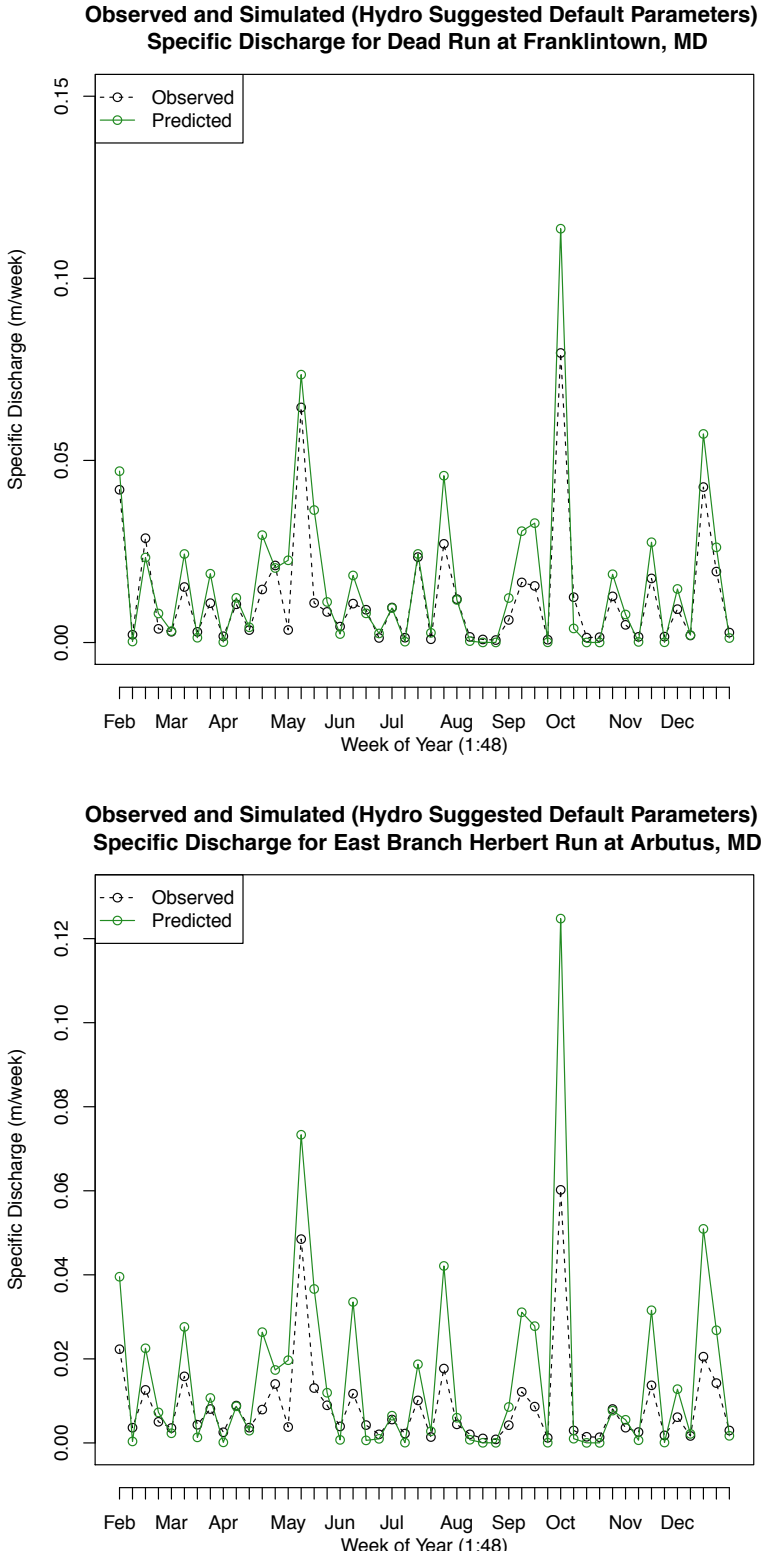


Figure A1.7- Plot of Weekly Aggregated Streamflow with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from a Run with Hydro Suggested Default Parameter Values (Predicted) at Dead Run (top) and Herbert Run (bottom).

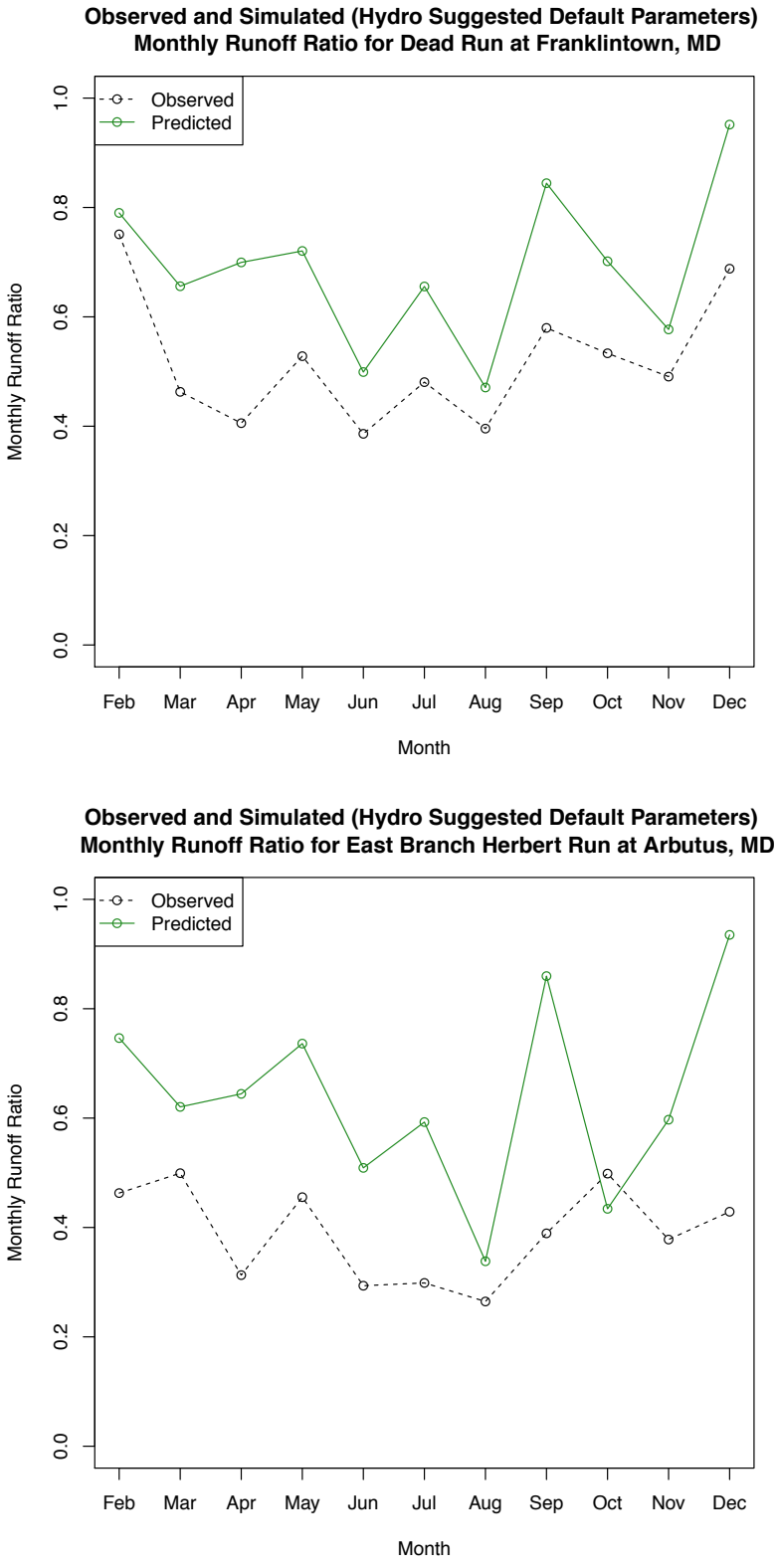


Figure A1.8- Plot of Monthly Runoff Ratios with USGS Daily Average Streamflow Data (Observed) Hydro Simulated Streamflow from a Run With the Hydro Suggested Default Parameter Values (Predicted) at Dead Run (top) and Herbert Run (bottom).

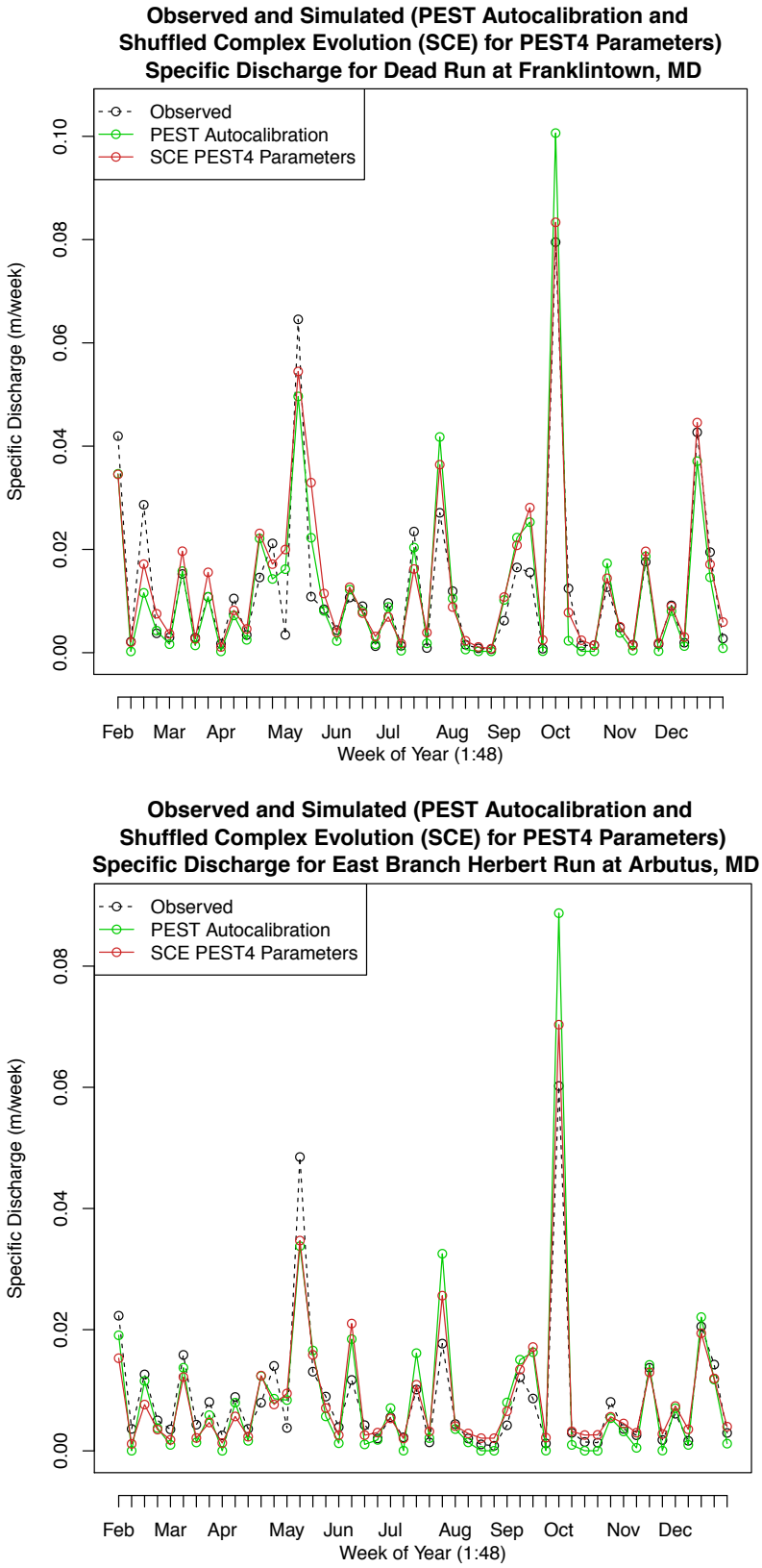
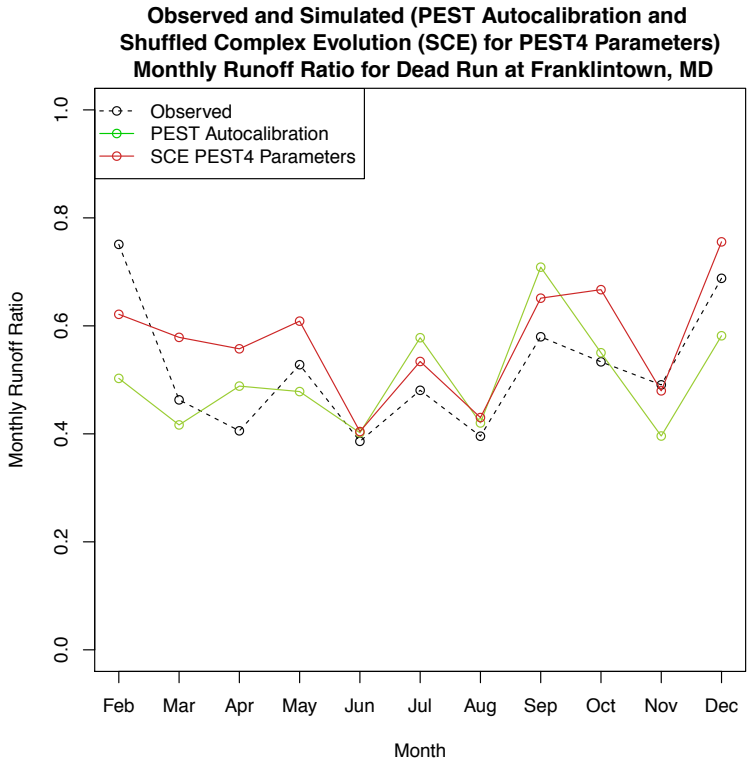


Figure A1.9- Plot of Weekly Aggregated Streamflow with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after PEST Autocalibration (PEST) and Shuffled Complex Evolution Calibration on the PEST4 Parameters (SCE PEST4) with an Objective Function of the Weekly Residual Sum of Squares at Dead Run (top) and Herbert Run (bottom).



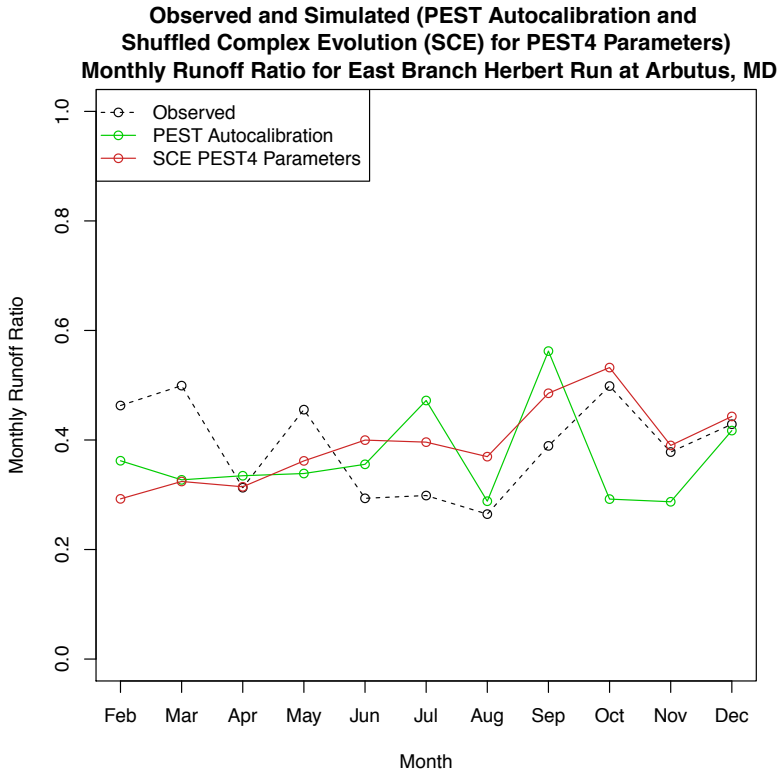


Figure A1.10- Plot of Monthly Runoff Ratios with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after PEST Autocalibration (PEST) and Shuffled Complex Evolution Calibration on the PEST4 Parameters (SCE PEST4) with an Objective Function of the Weekly Residual Sum of Squares at Dead Run (top) and Herbert Run (bottom).

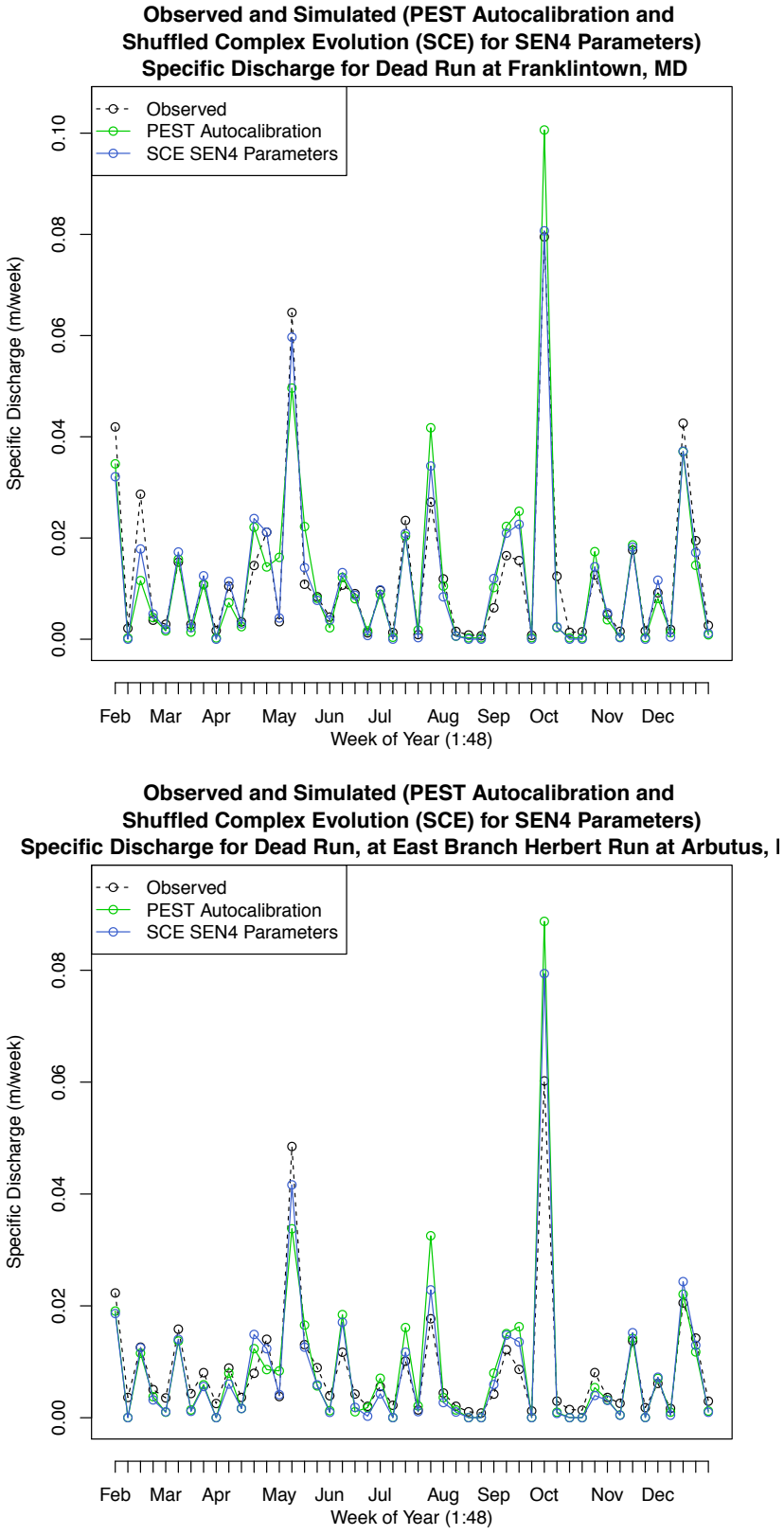
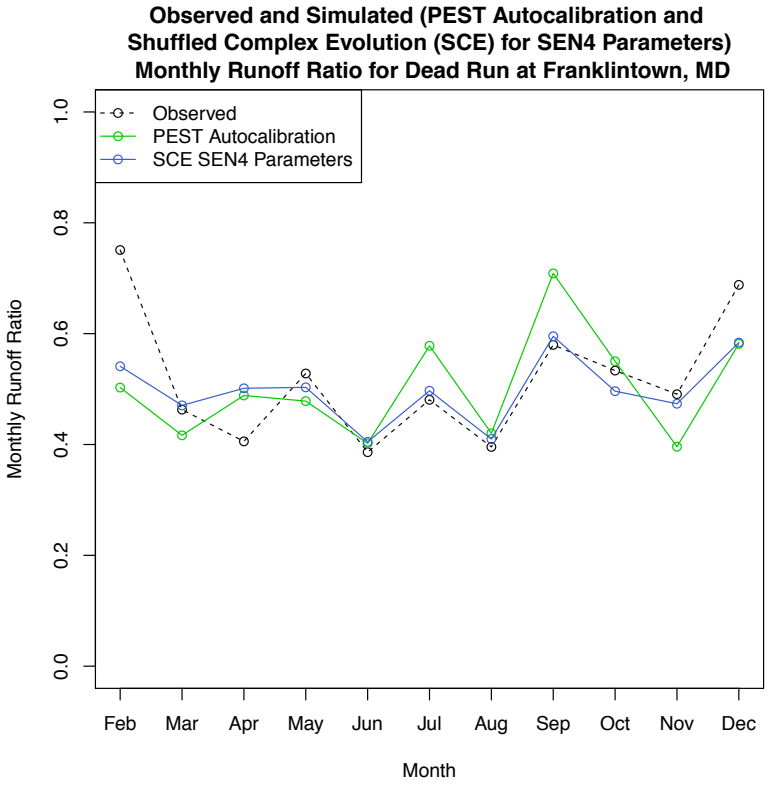


Figure A1.11- Plot of Weekly Aggregated Streamflow with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after PEST Autocalibration (PEST) and Shuffled Complex Evolution Calibration on the SEN4 Parameters (SCE SEN4) with an Objective Function of the Weekly Residual Sum of Squares at Dead Run (top) and Herbert Run (bottom).



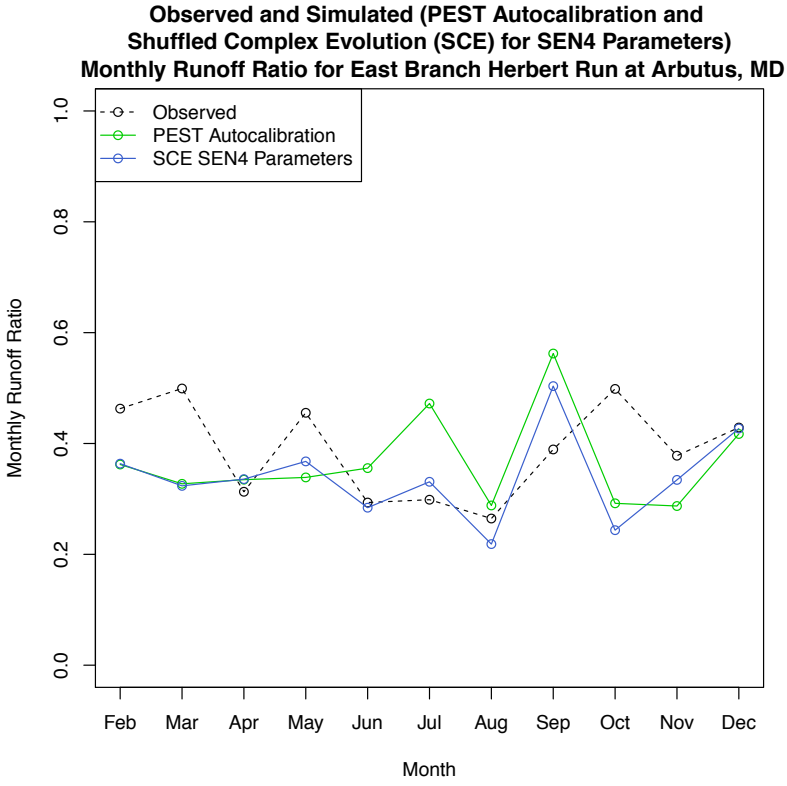
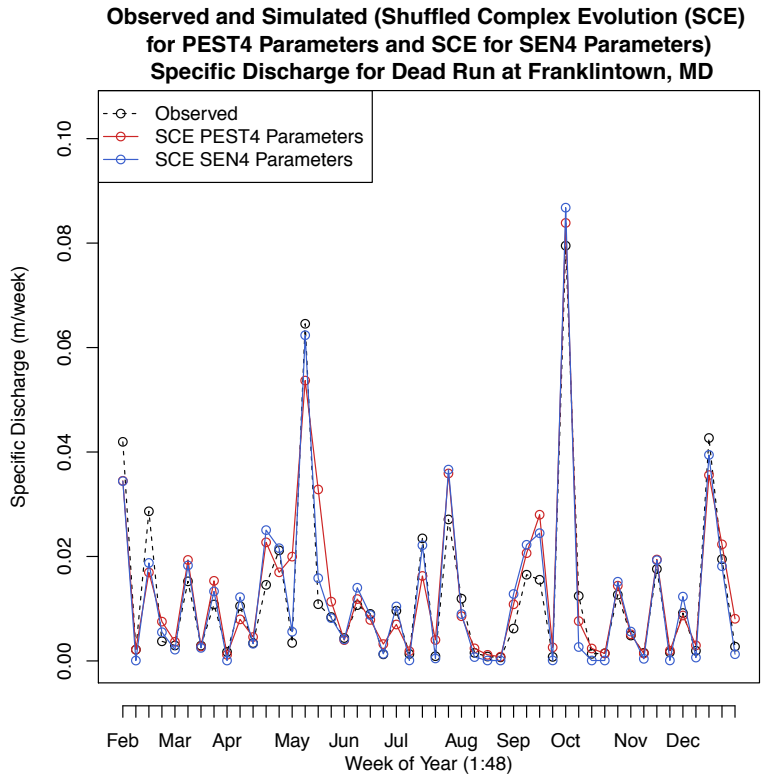


Figure A1.12- Plot of Monthly Runoff Ratios with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after PEST Autocalibration (PEST) and Shuffled Complex Evolution Calibration on the SEN4 Parameters (SCE SEN4) with an Objective Function of the Weekly Residual Sum of Squares at Dead Run (top) and Herbert Run (bottom).



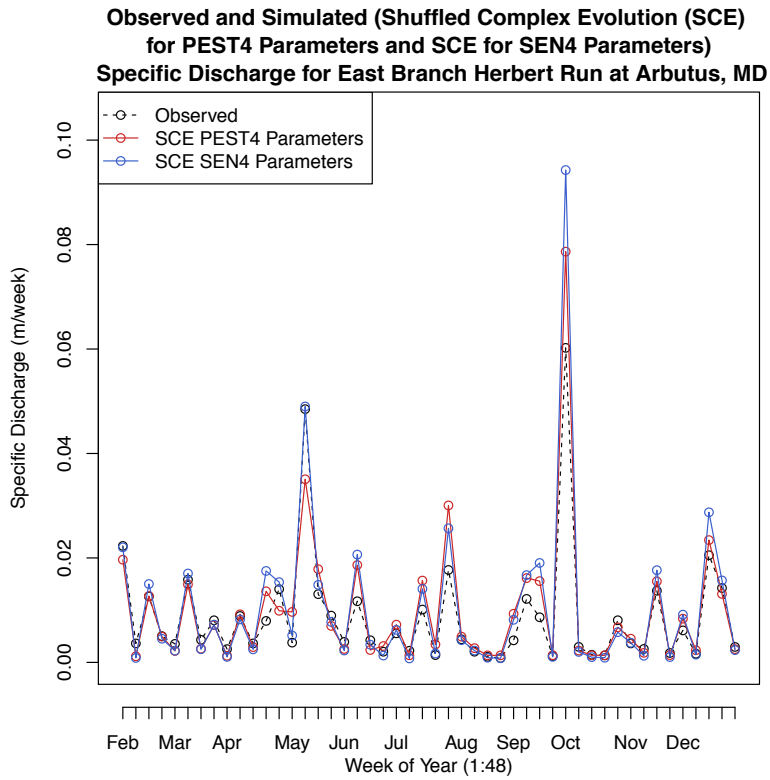
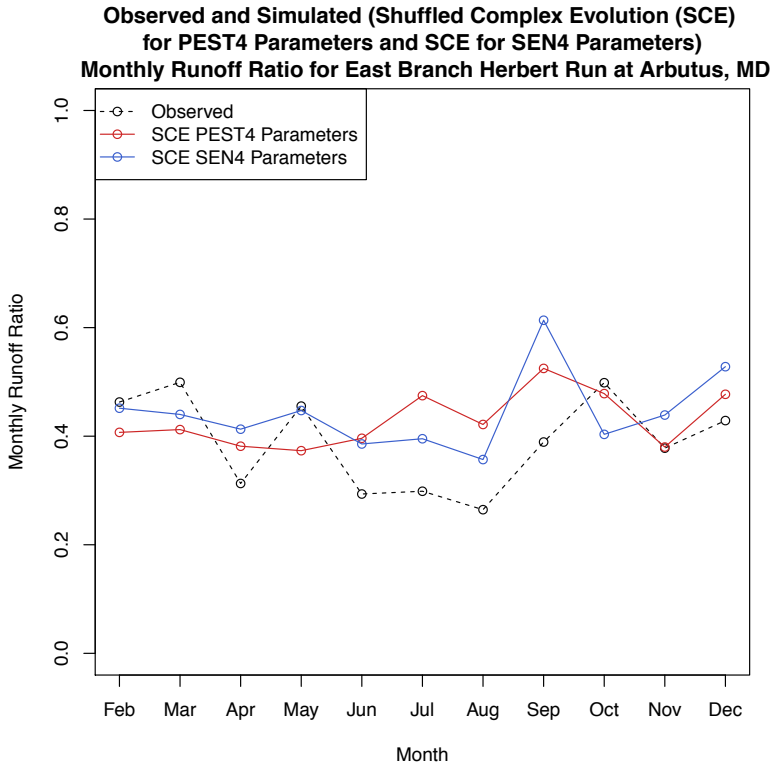


Figure A1.13- Plot of Weekly Aggregated Streamflow with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after Shuffled Complex Evolution Calibrations on the PEST4 (SCE PEST4) SEN4 Parameters (SCE SEN4) with an Objective Function of the Monthly Runoff Ratios at Dead Run (top) and Herbert Run (bottom).



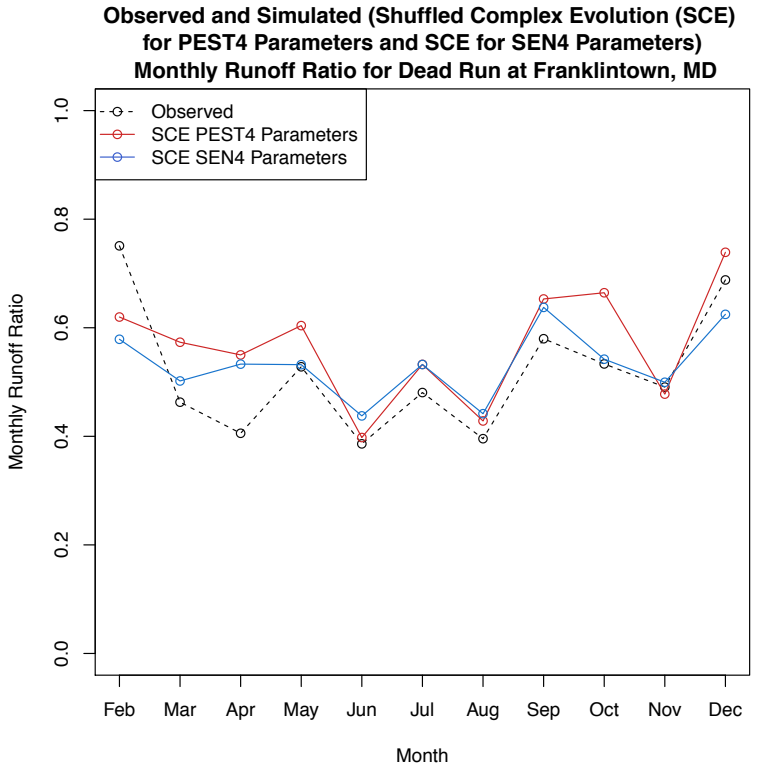


Figure A1.14- Plot of Monthly Runoff Ratios with USGS Daily Average Streamflow Data (Observed) and Hydro Simulated Streamflow from Runs of Hydro after Shuffled Complex Evolution Calibrations on the PEST4 (SCE PEST4) SEN4 Parameters (SCE SEN4) with an Objective Function of the Monthly Runoff Ratios at Dead Run (top) and Herbert Run (bottom).

Appendix 2: Additional Figures

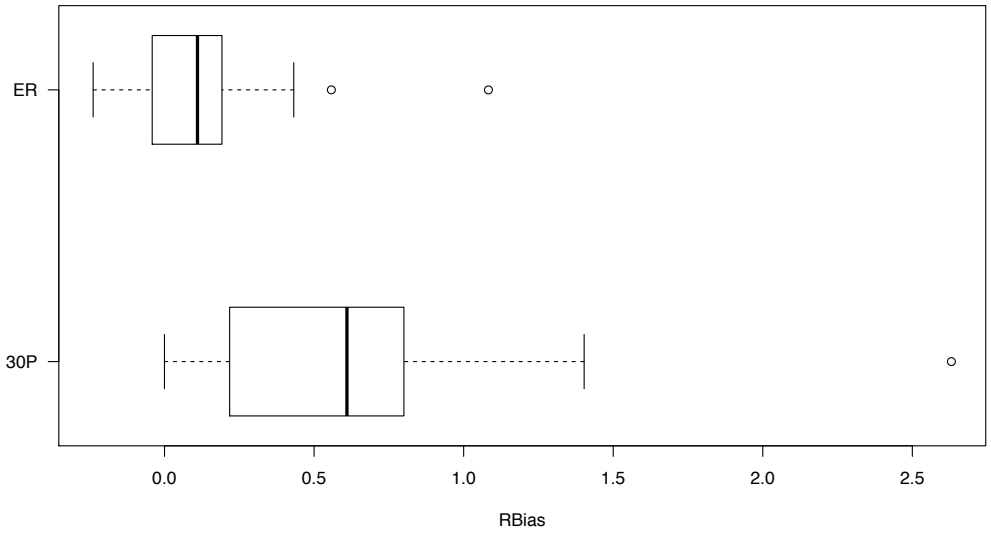


Figure A.2.1- Relative Bias (RBias) for non-overlapping 14-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

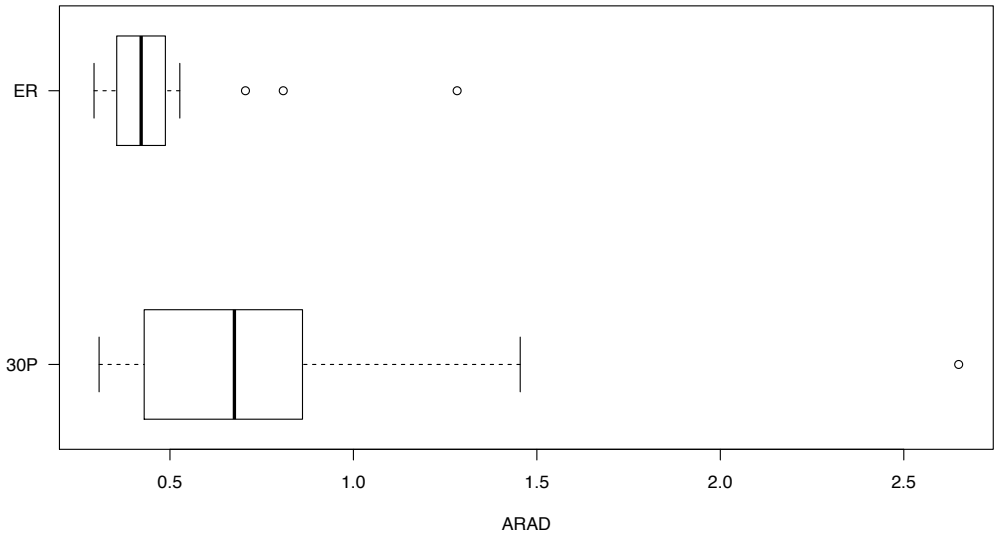


Figure A.2.2- Absolute Relative Absolute Difference (ARAD) for non-overlapping 14-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

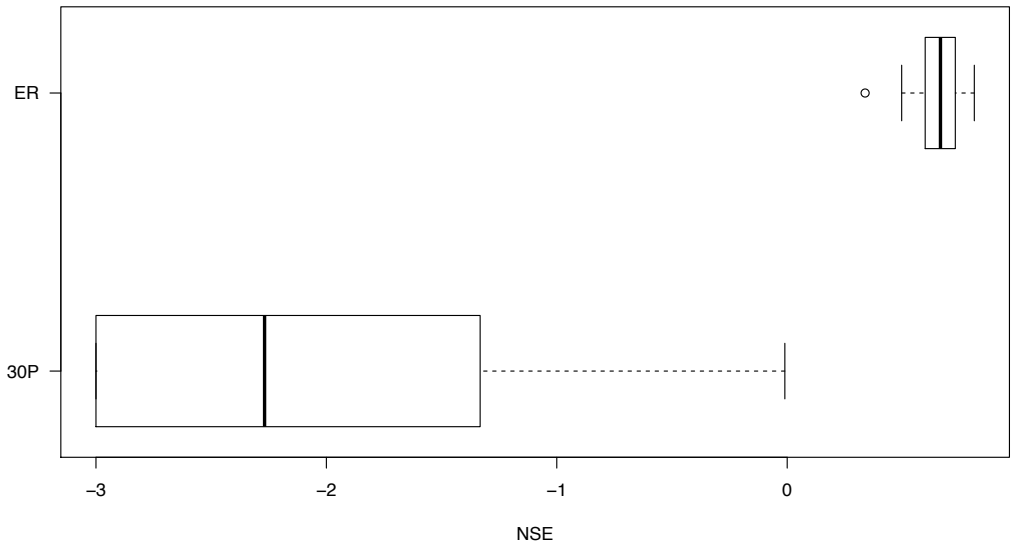


Figure A.2.3- Real Space Nash-Sutcliffe Efficiency (NSE) for non-overlapping 14-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

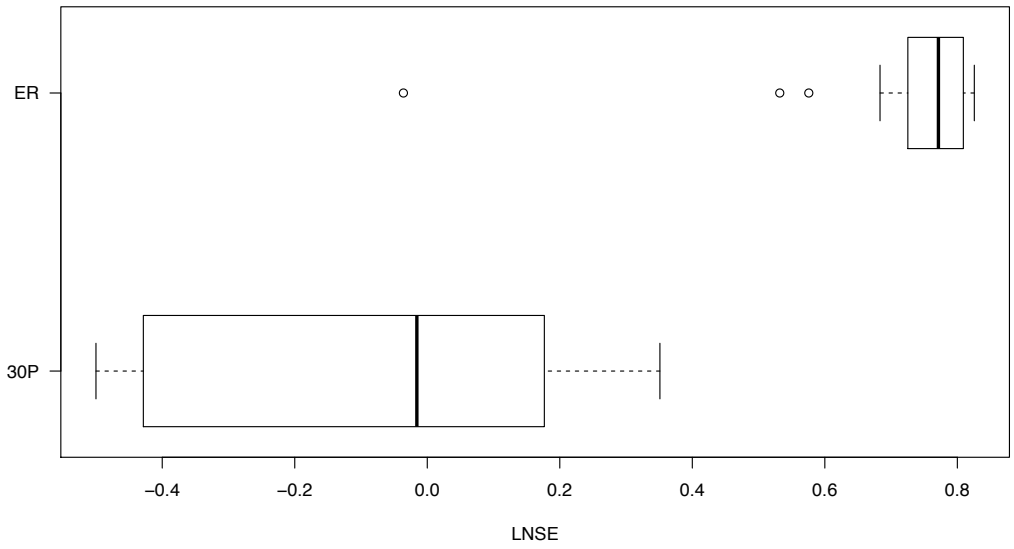


Figure A.2.4- Log-Space Nash Sutcliffe Efficiency (LNSE) for non-overlapping 14-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

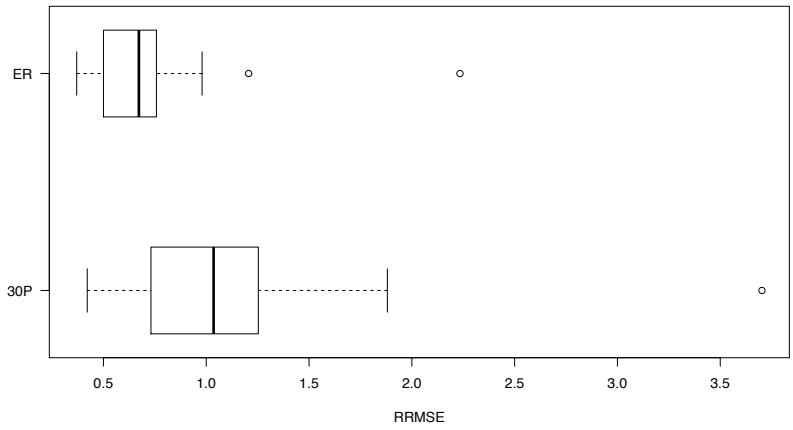


Figure A.2.5- Relative Root Mean Square Error (RMSE) for non-overlapping 14-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

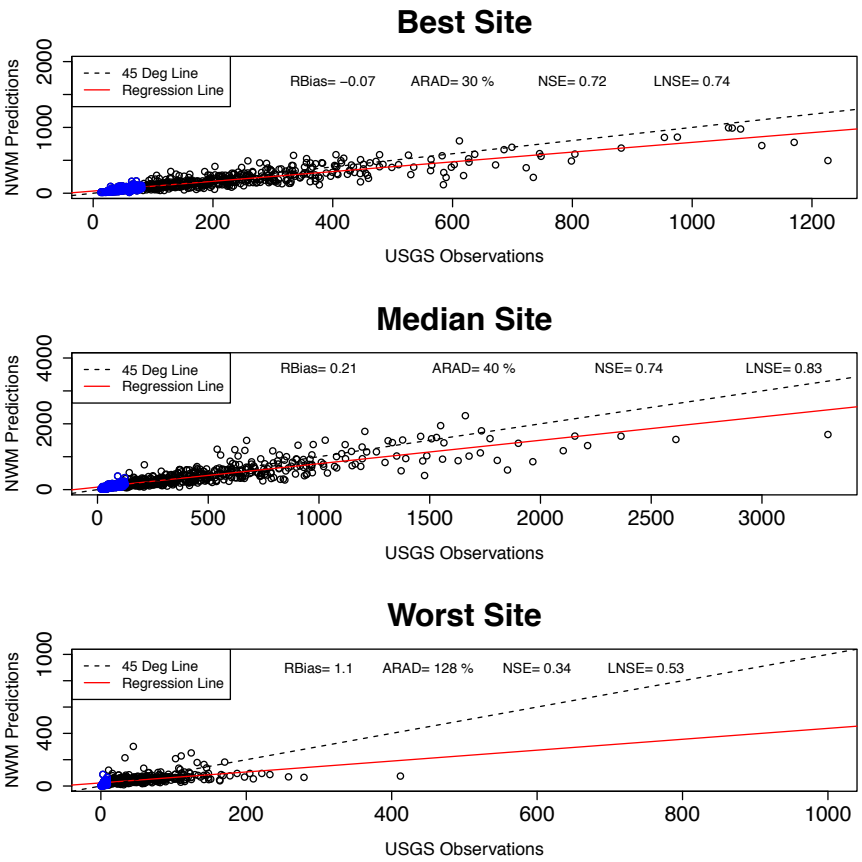


Figure A.2.6- Plots of NWM 14-day average streamflow versus USGS observed 14-day average streamflow at site with the lowest ARAD, median ARAD, and highest ARAD

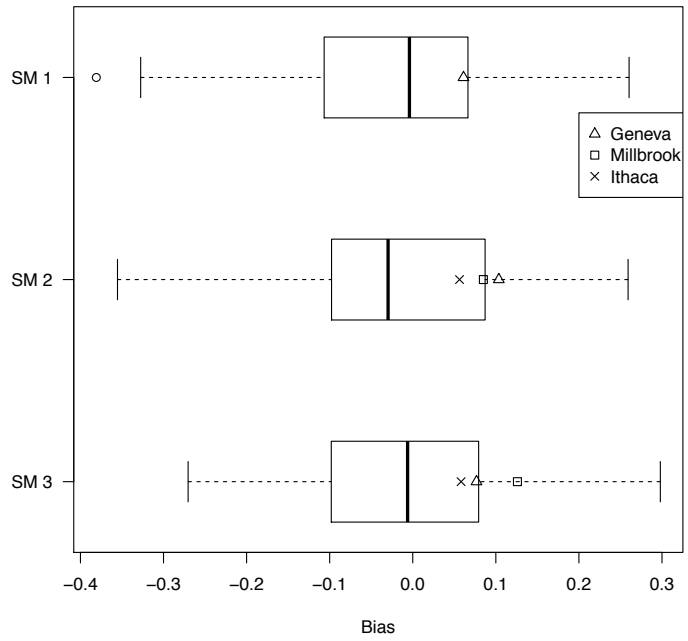


Figure A.2.7- Bias in the SM1, SM2, and SM3 Soil Moisture Profile for a 14-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

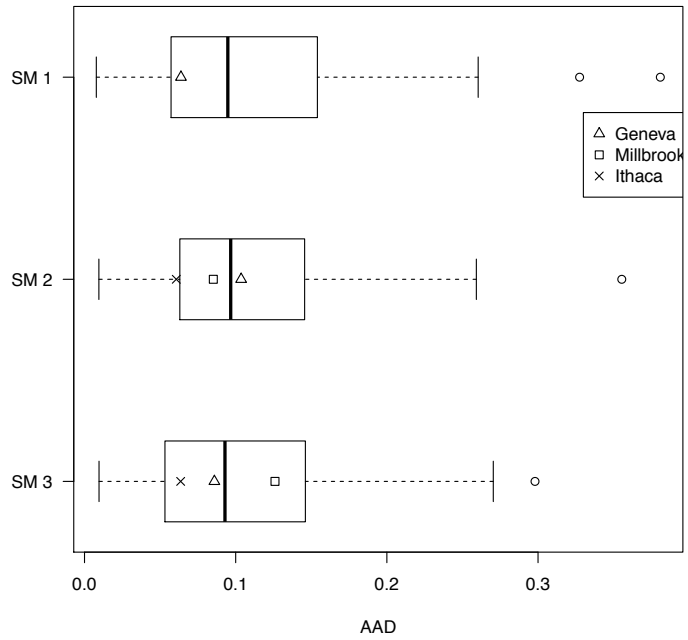


Figure A.2.8- AAD in the SM1, SM2, and SM3 Soil Moisture Profile for a 14-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

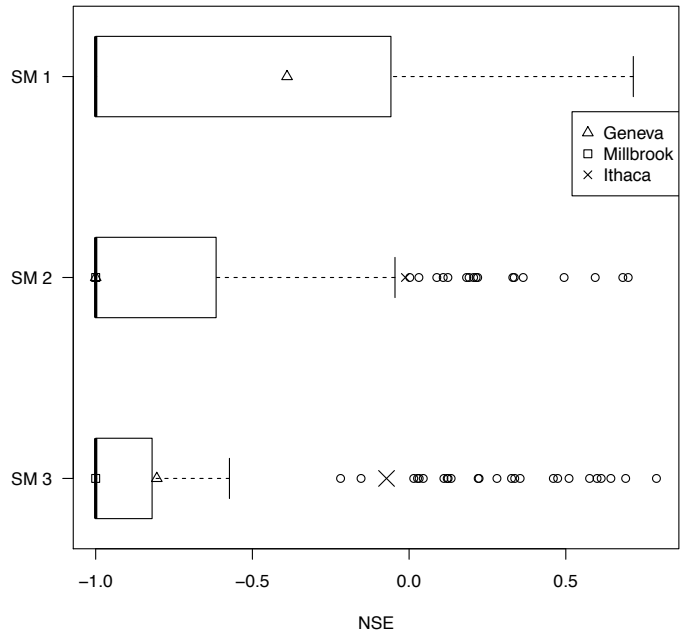


Figure A.2.9- NSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 14-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

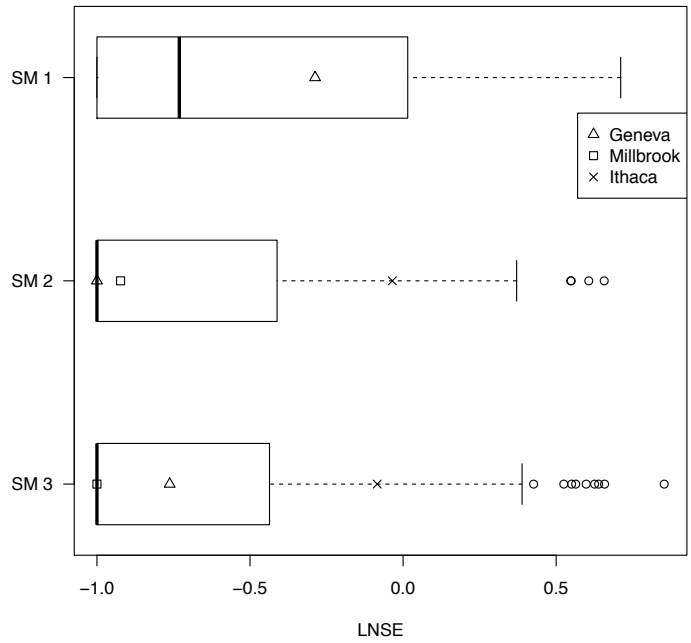


Figure A.2.10- LNSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 14-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

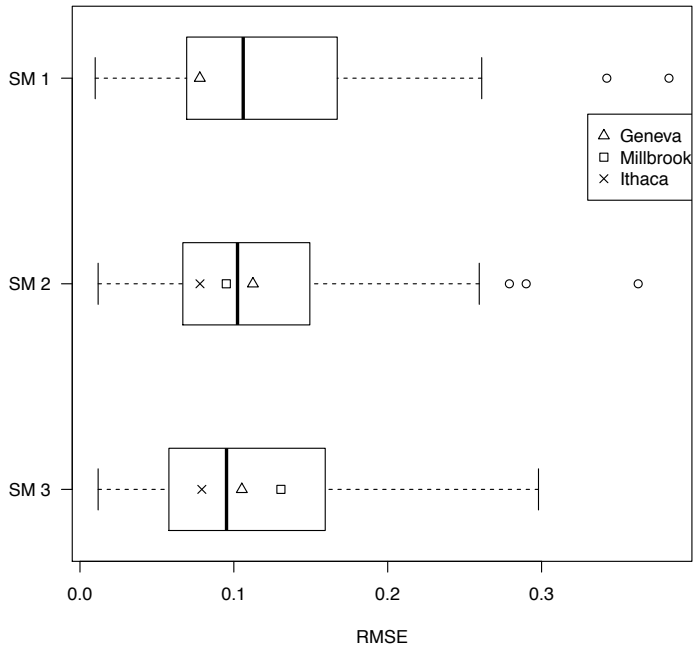


Figure A.2.11- RMSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 14-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

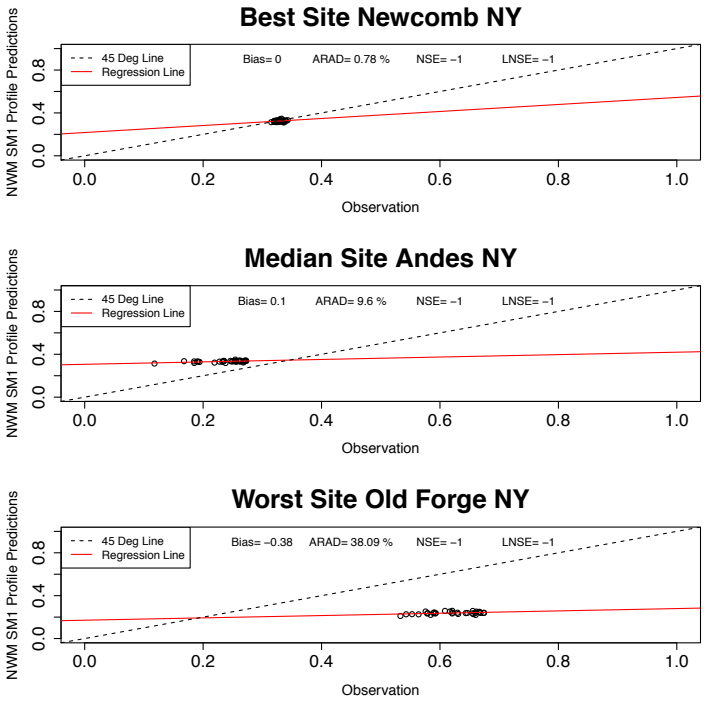


Figure A.2.12- Plots of NWM 14-day average SM1 profile soil moisture versus Mesonet observed 14-day average soil moisture at site with the lowest ARAD, median ARAD, and highest ARAD.

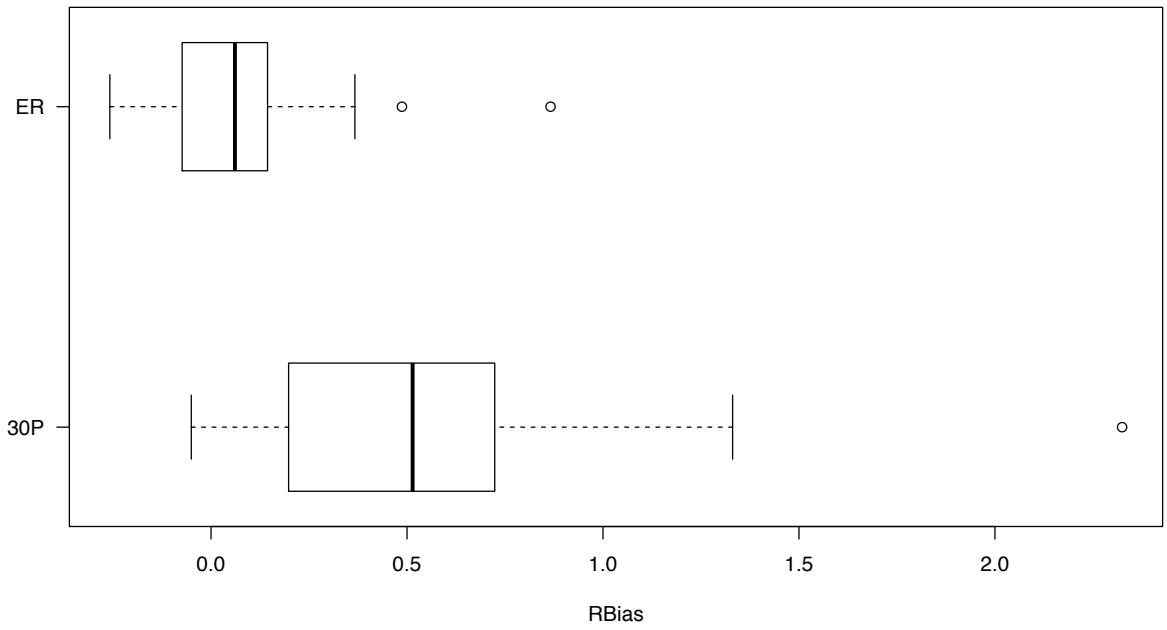


Figure A.2.13- Relative Bias (RBias) for non-overlapping 28-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

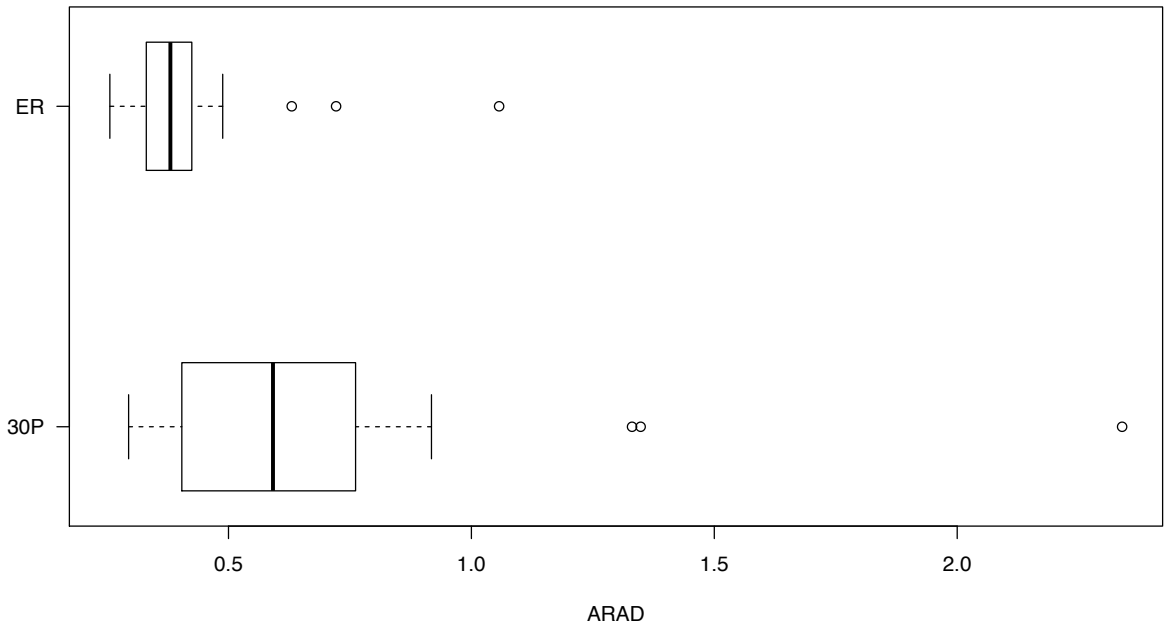


Figure A.2.14- Absolute Relative Absolute Difference (ARAD) for non-overlapping 28-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

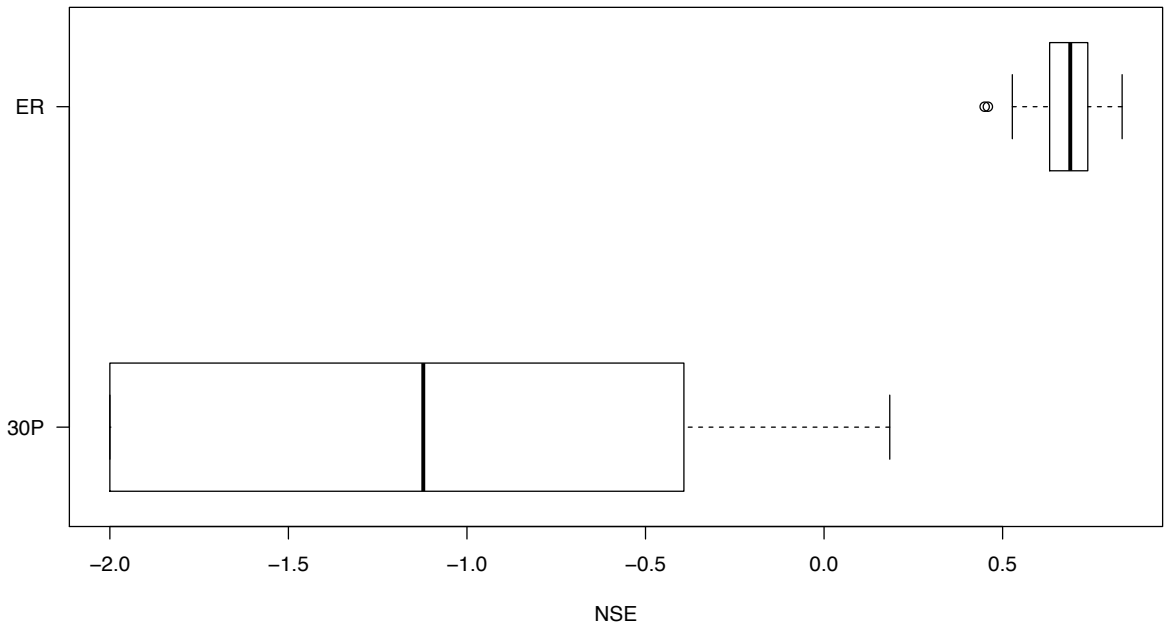


Figure A.2.15- Real Space Nash-Sutcliffe Efficiency (NSE) for non-overlapping 28-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

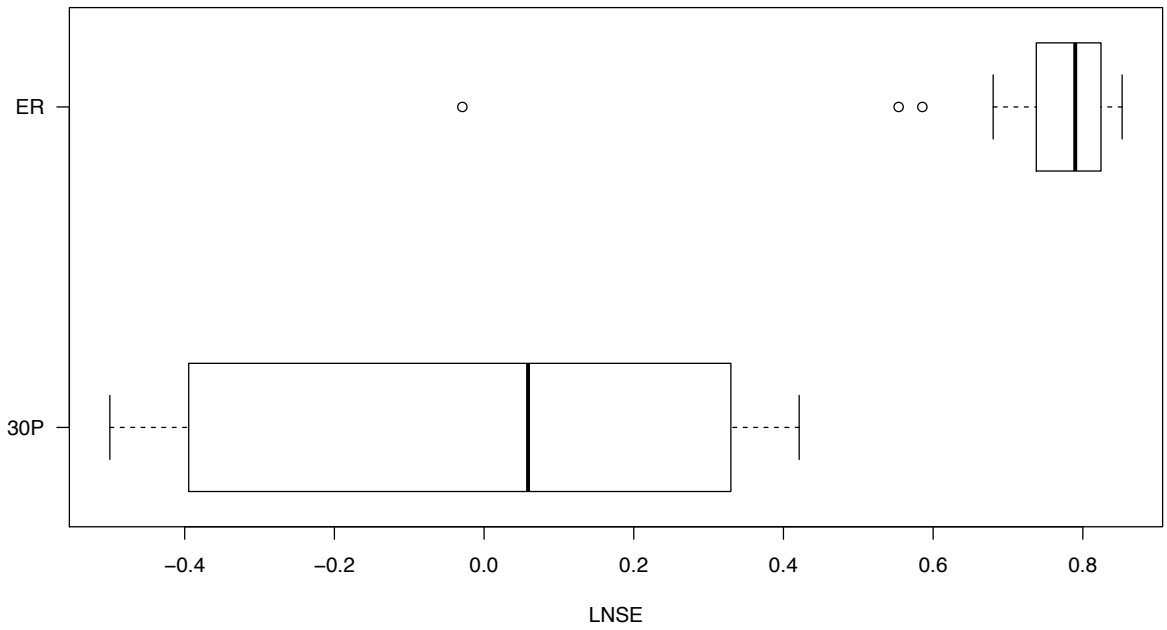


Figure A.2.16- Log-Space Nash Sutcliffe Efficiency (LNSE) for non-overlapping 28-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

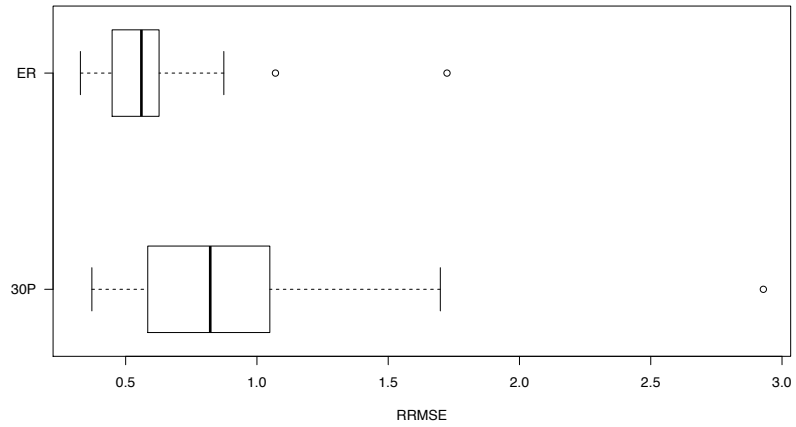


Figure A.2.17- Relative Root Mean Square Error (RMSE) for non-overlapping 28-day averages at the 28 USGS gauging stations for entire record and for less than the 30th percentile.

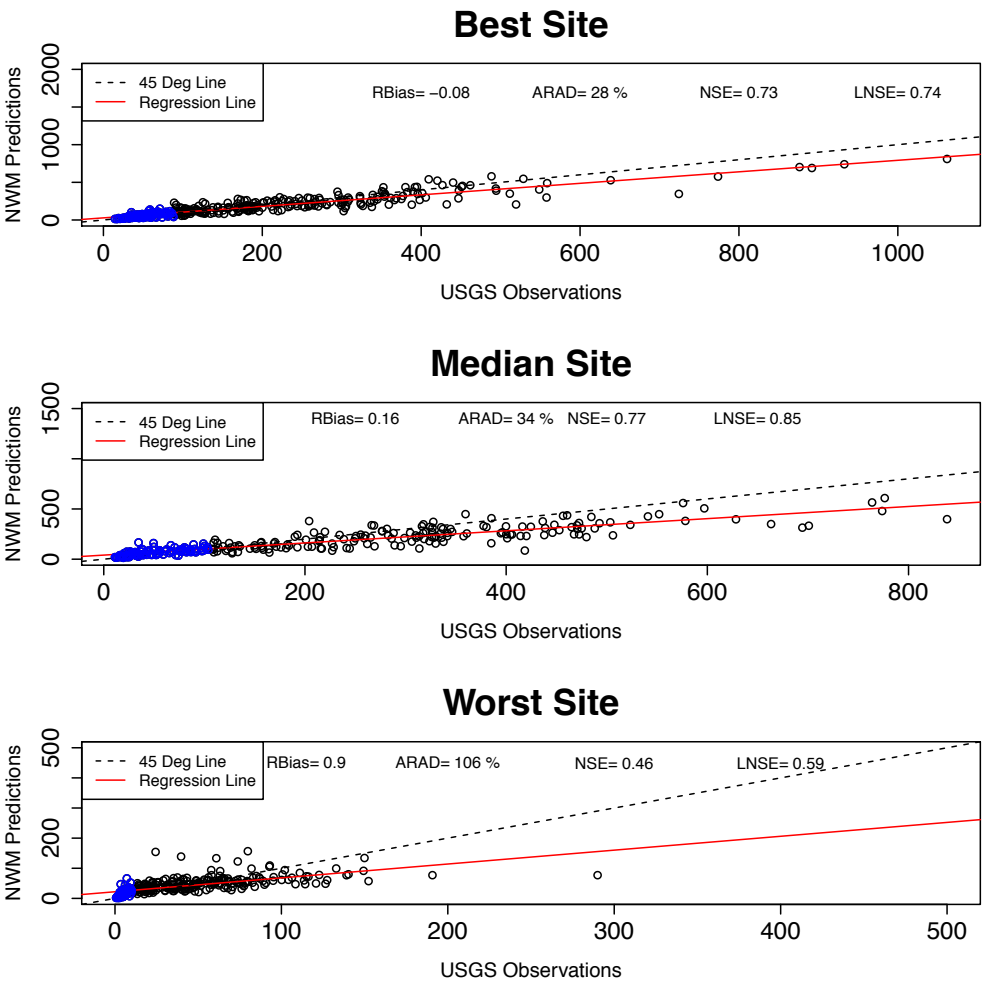


Figure A.2.18- Plots of NWM 28-day average streamflow versus USGS observed 28-day average streamflow at site with the lowest ARAD, median ARAD, and highest ARAD

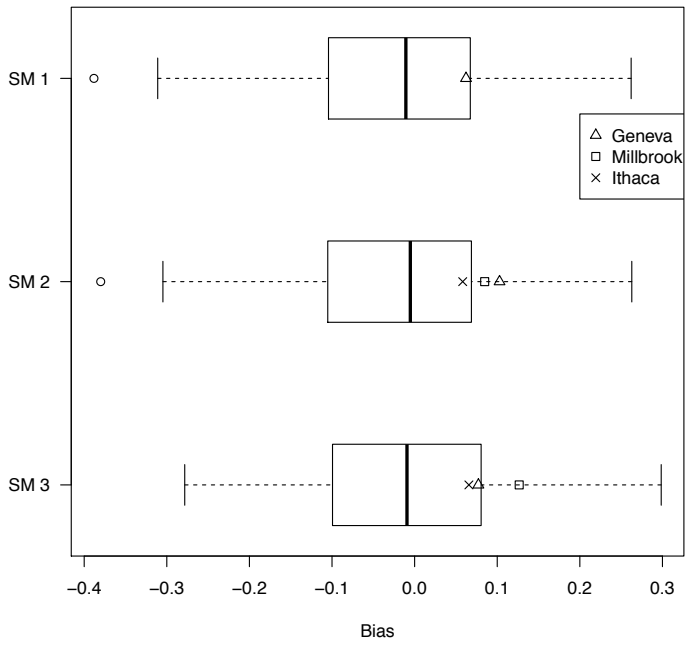


Figure A.2.19- Bias in the SM1, SM2, and SM3 Soil Moisture Profile for a 28-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

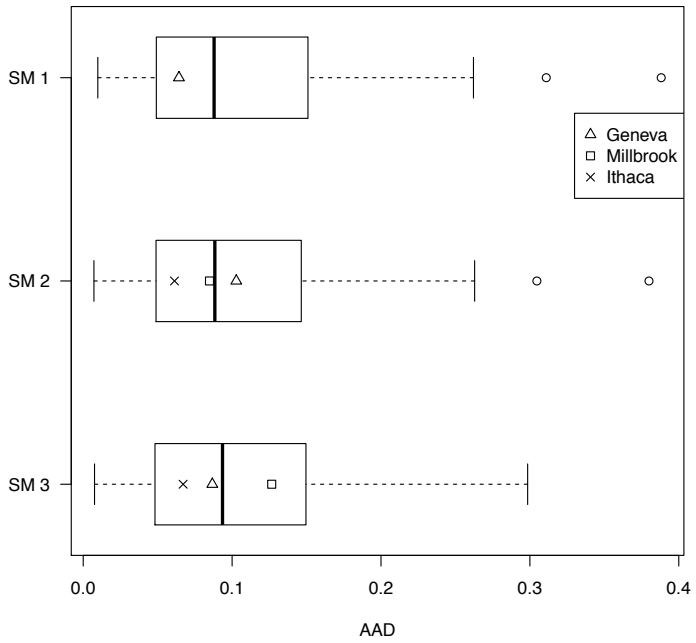


Figure A.2.20- AAD in the SM1, SM2, and SM3 Soil Moisture Profile for a 28-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

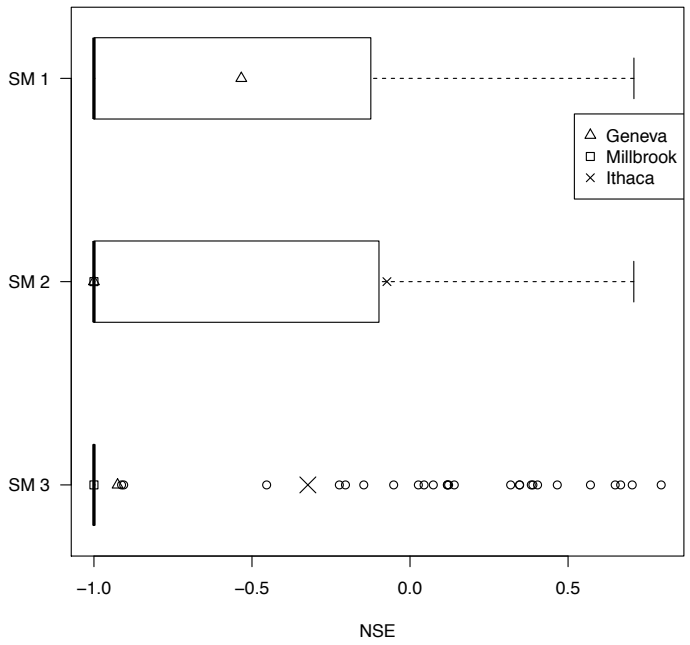


Figure A.2.21- NSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 28-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

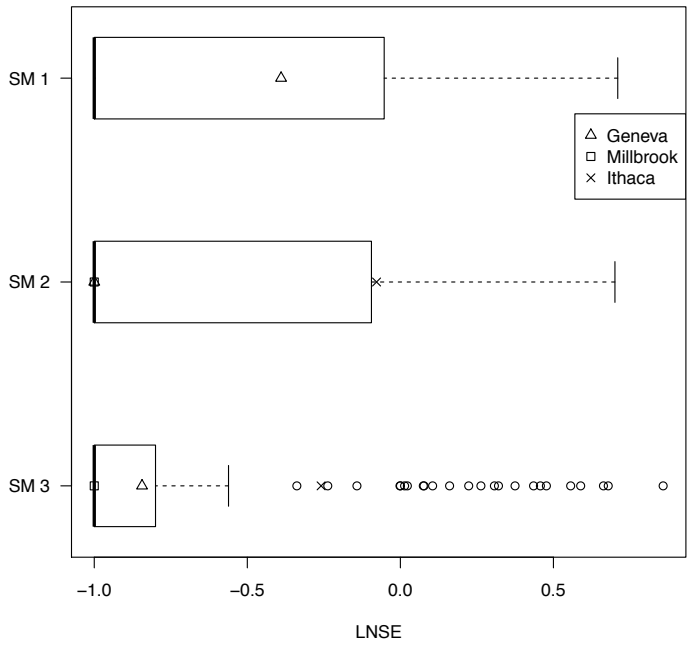


Figure A.2.22- LNSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 28-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

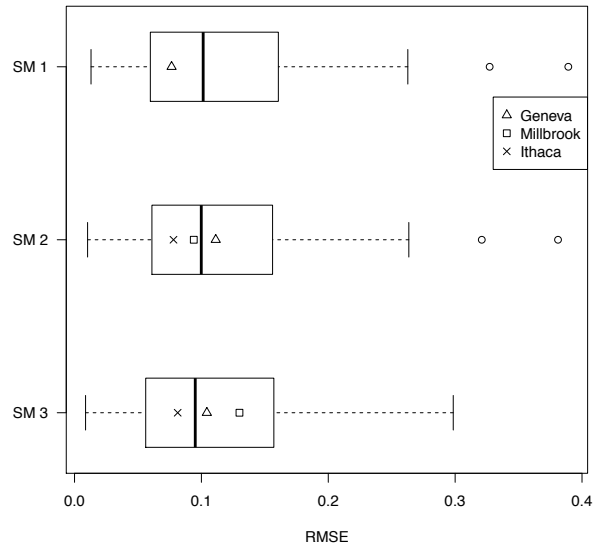


Figure A.2.23- RMSE in the SM1, SM2, and SM3 Soil Moisture Profile for a 28-day non-overlapping averaging period of Daily Aggregated Sub-Hourly Soil Moisture Observations.

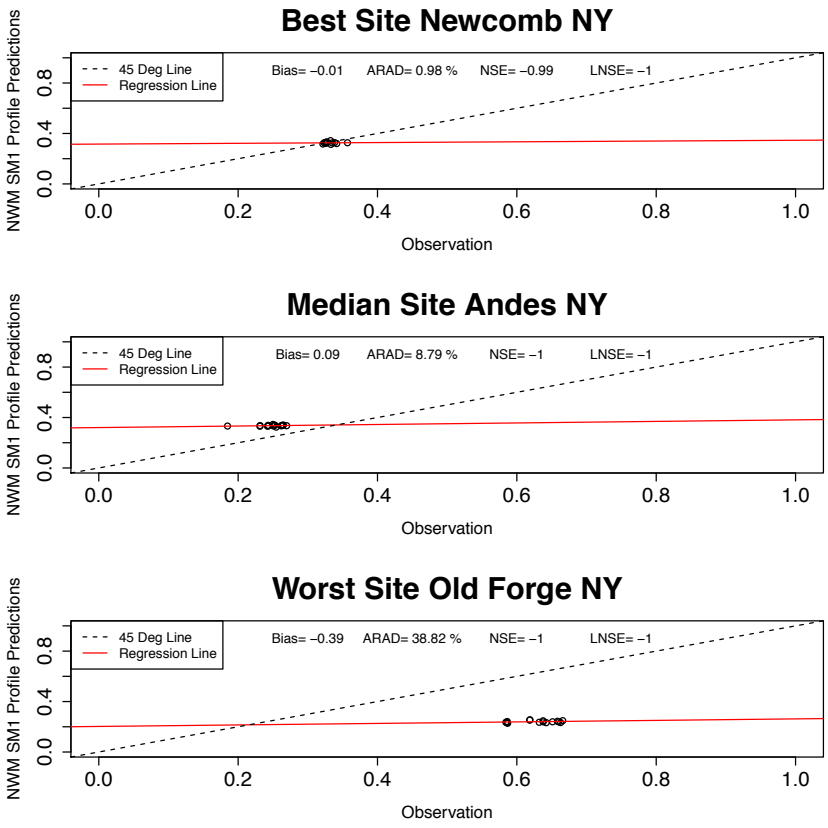


Figure A.2.24- Plots of NWM 28-day average SM1 profile soil moisture versus Mesonet observed 28-day average soil moisture at site with the lowest ARAD, median ARAD, and highest ARAD.

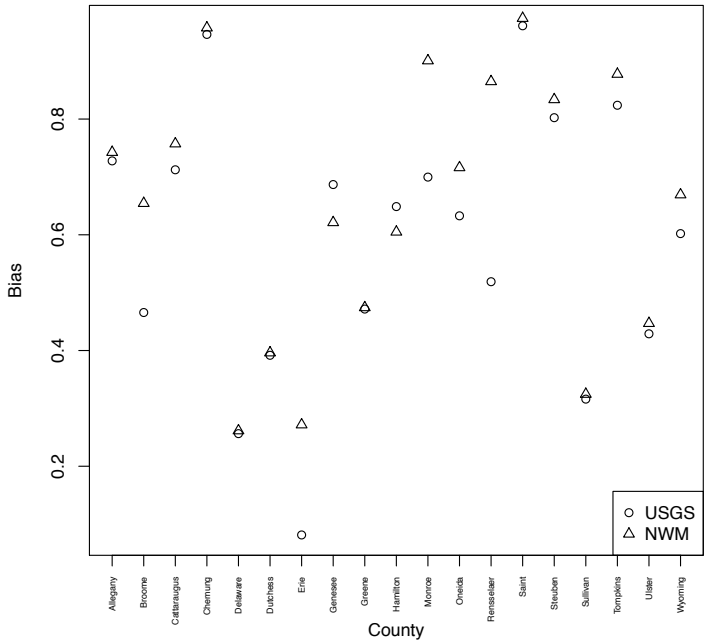


Figure A.2.25- Bias for 7-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

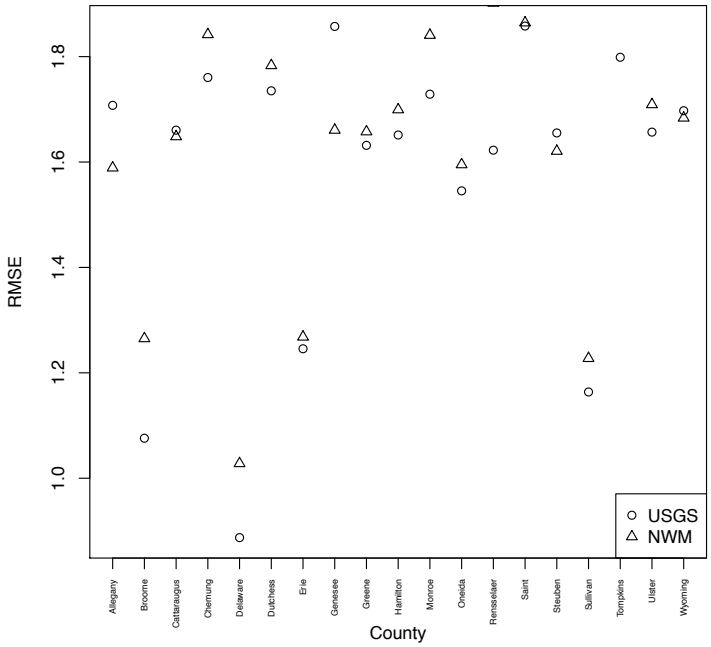


Figure A.2.26- RMSE for 7-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

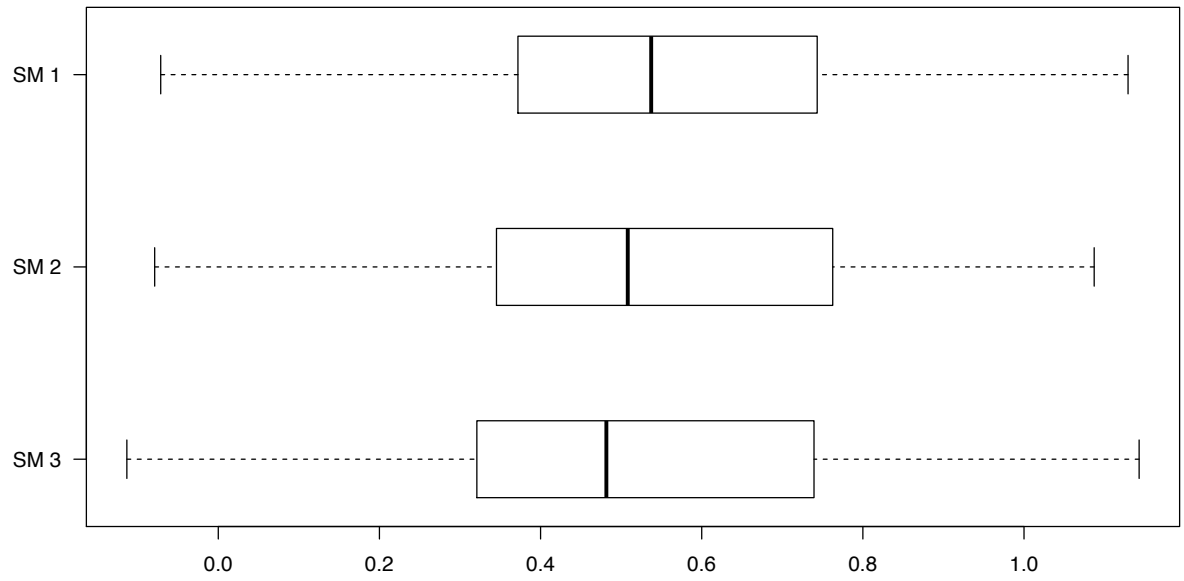


Figure A.2.27- Bias for 7-day average NWM soil moisture drought categories as compared to county USDM drought majority (M) categories.

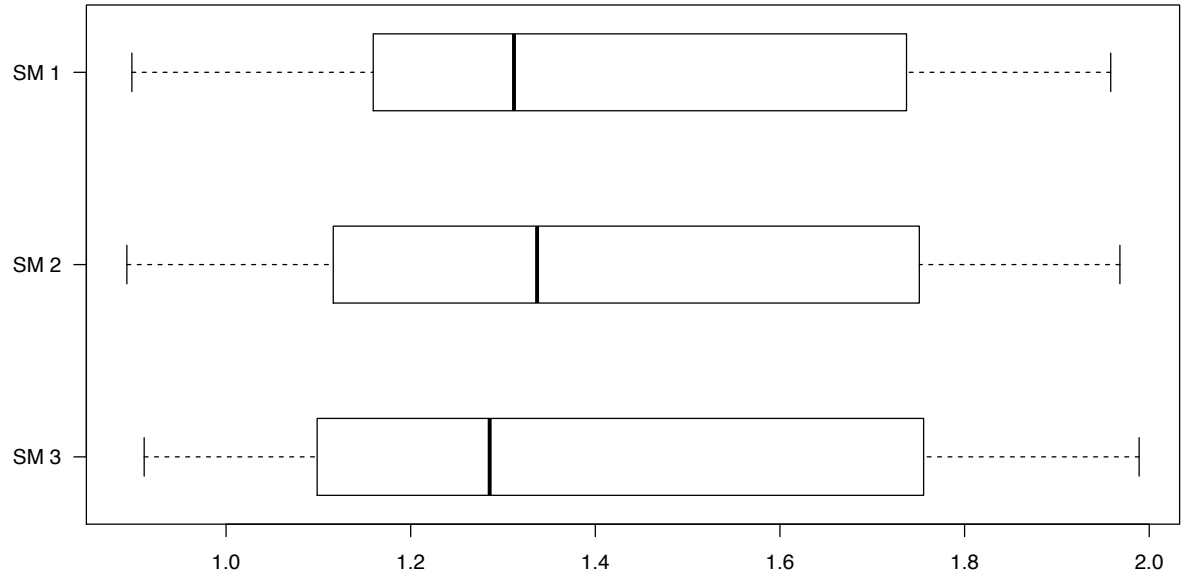


Figure A.2.28- RMSE for 7-day average NWM soil moisture drought categories as compared to county USDM drought (M) categories.

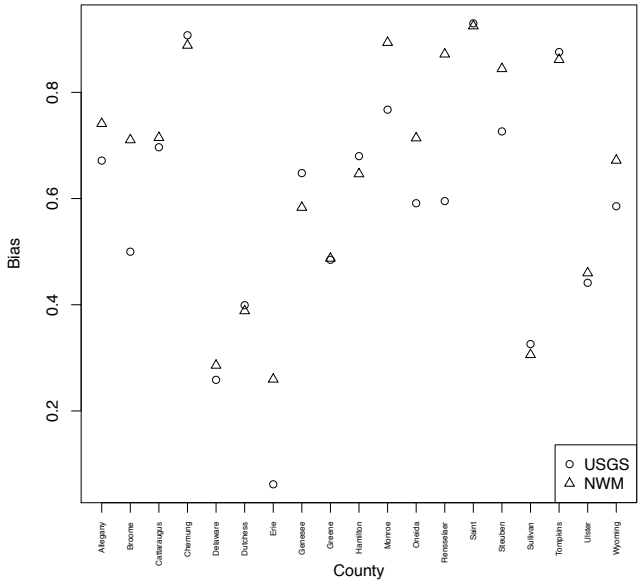


Figure A.2.29- Bias for 14-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

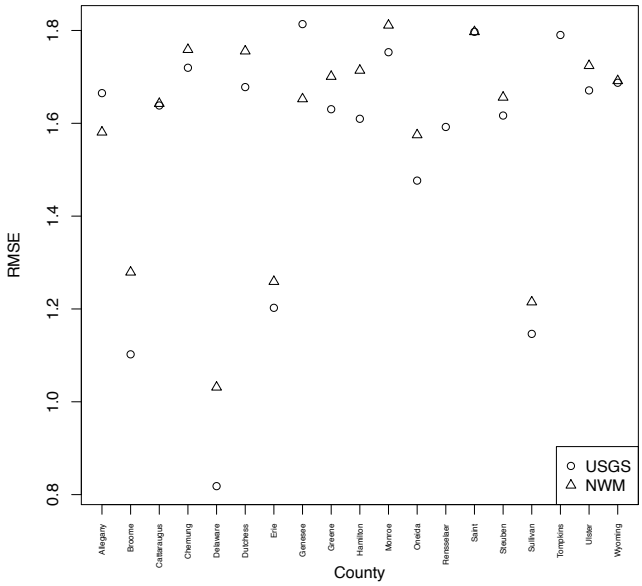


Figure A.2.30- RMSE for 14-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

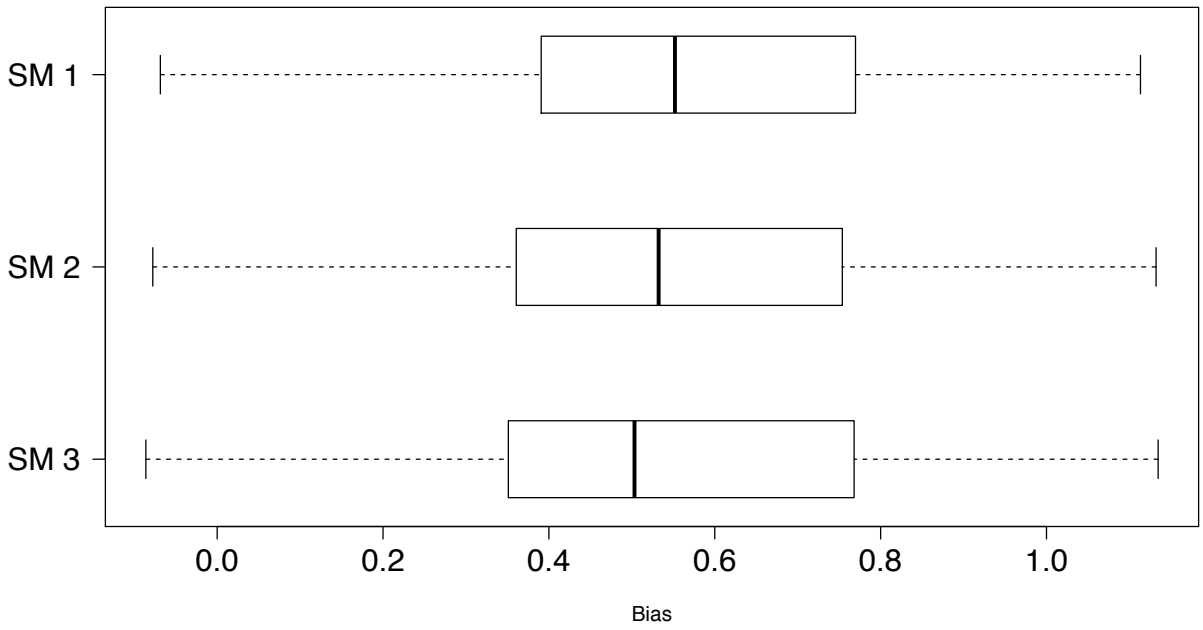


Figure A.2.31- Bias for 14-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories.

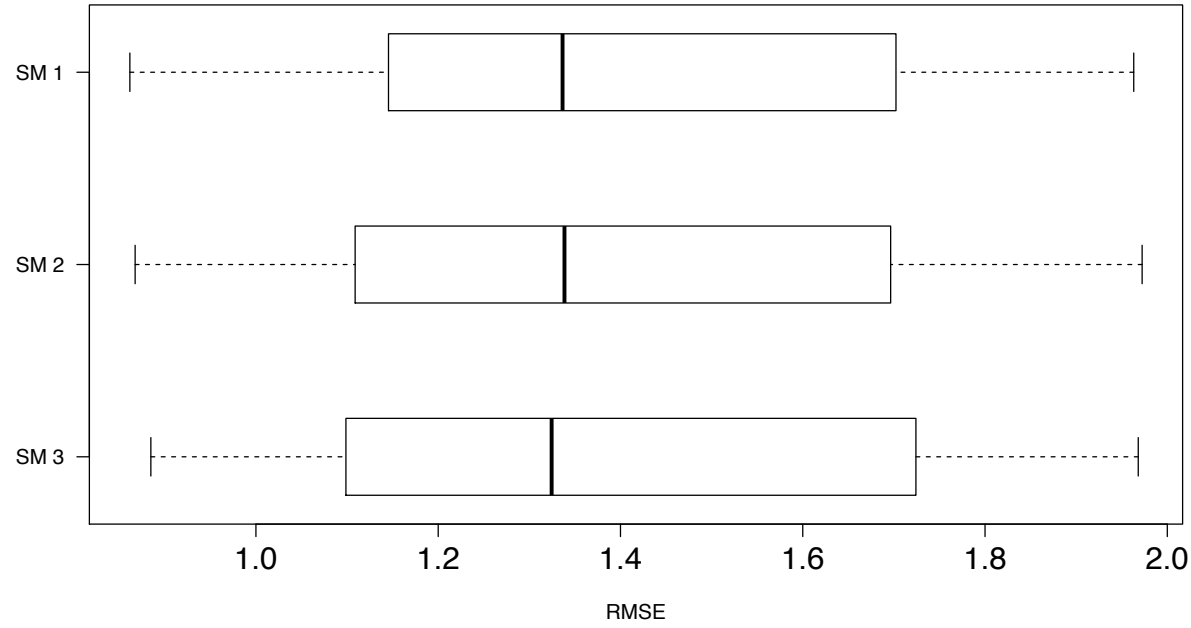


Figure A.2.32- RMSE for 14-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories

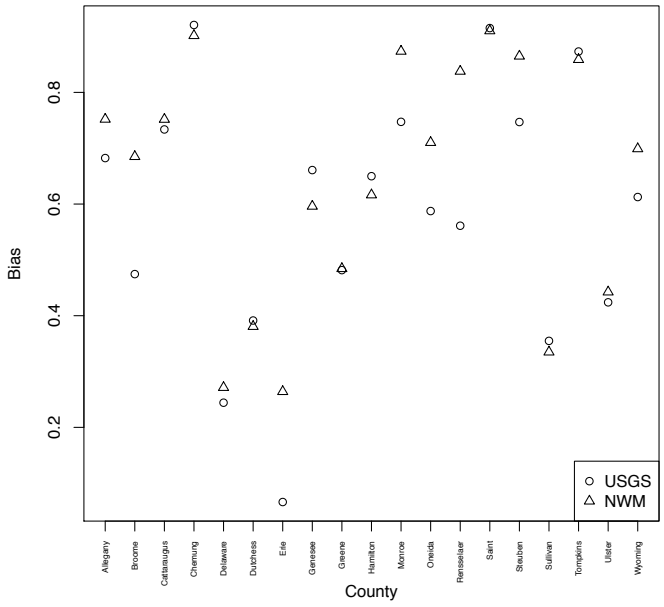


Figure A.2.33- Bias for 14-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

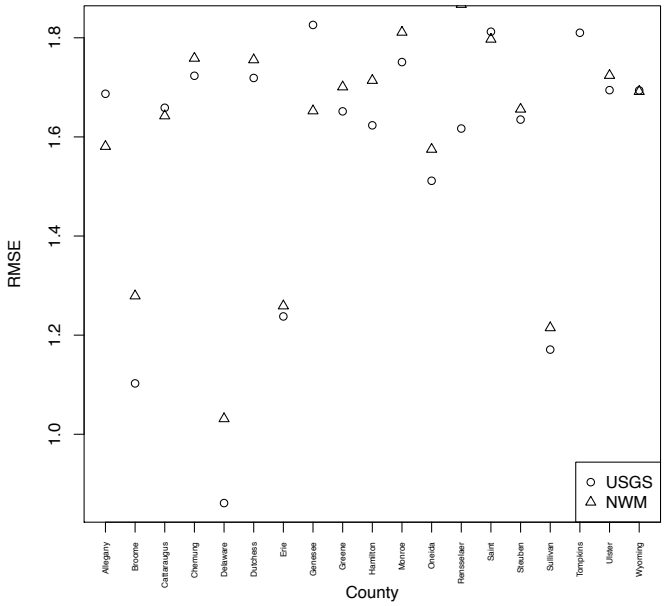


Figure A.2.34- RMSE for 14-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

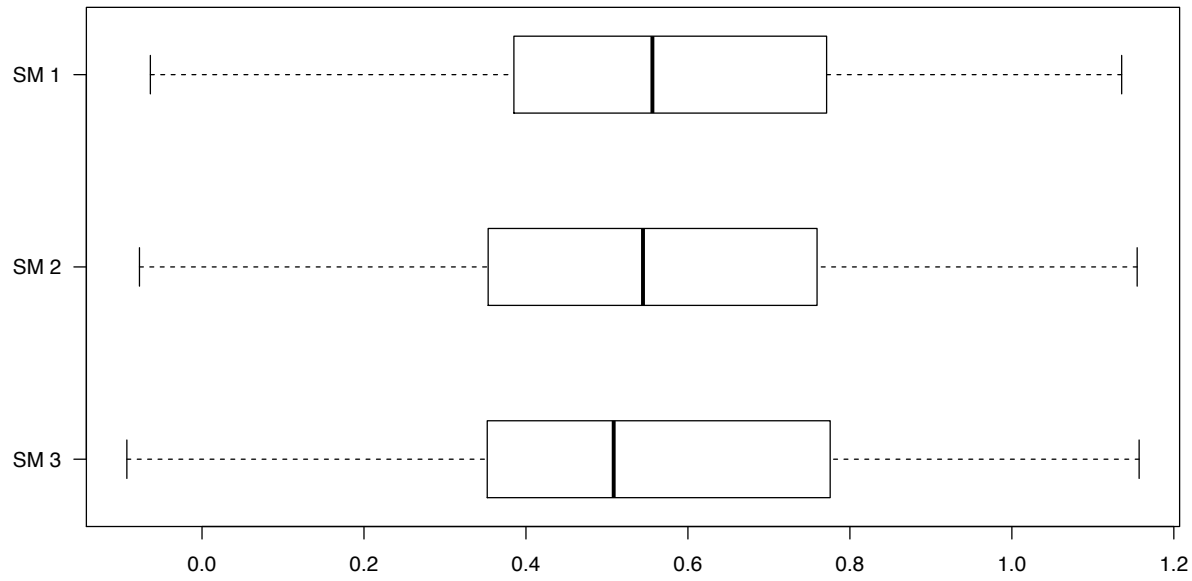


Figure A.2.35- Bias for 14-day average NWM soil moisture drought categories as compared to county USDM drought majority (M) categories.

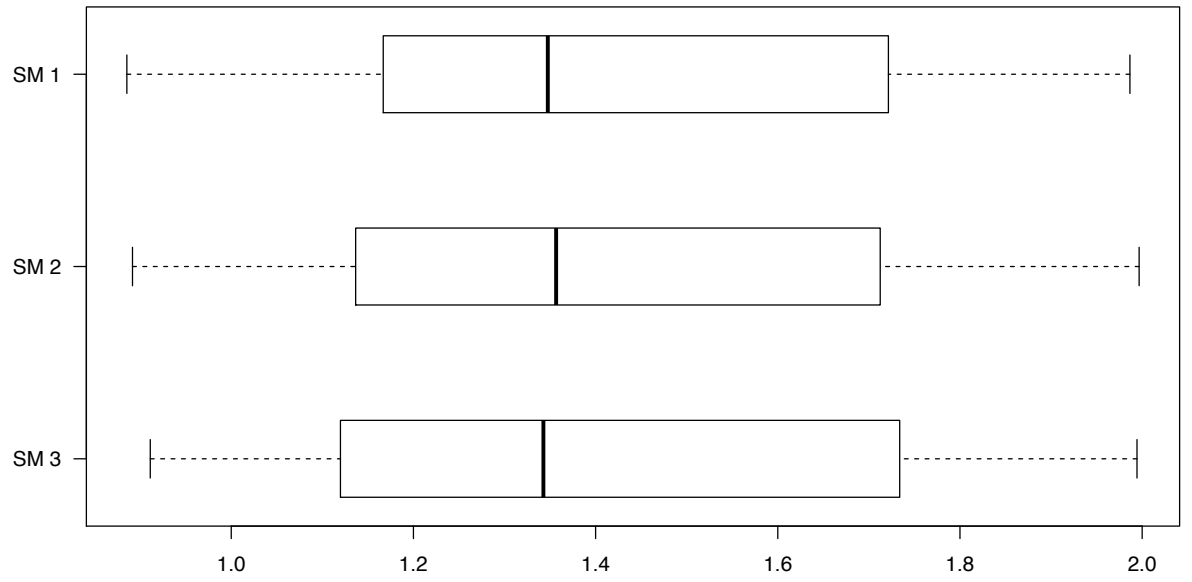


Figure A.2.36- RMSE for 14-day average NWM soil moisture drought categories as compared to county USDM drought majority (M) categories

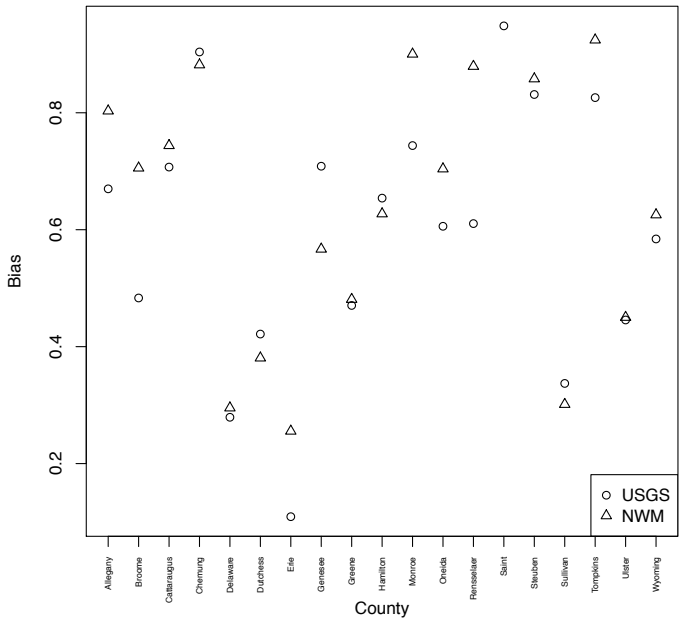


Figure A.2.37- Bias for 28-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

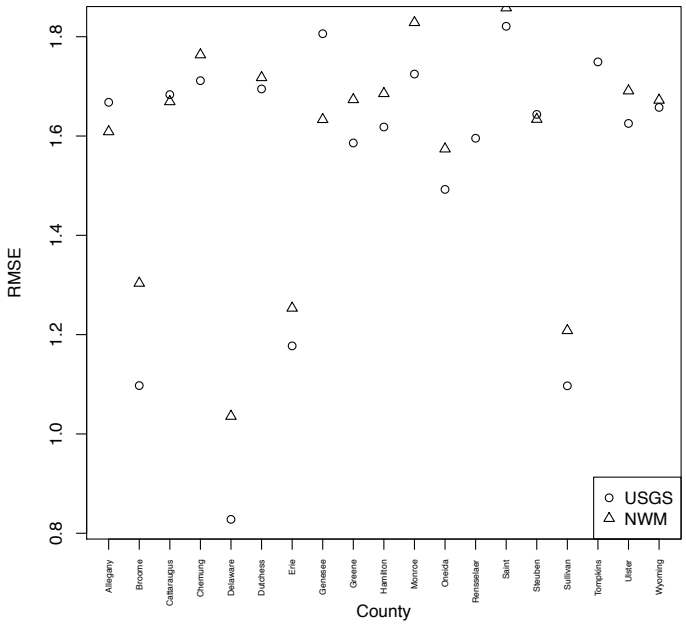


Figure A.2.38- RMSE for 28-day average USGS and NWM streamflow drought categories as compared to county USDM drought weighted average (WA) categories.

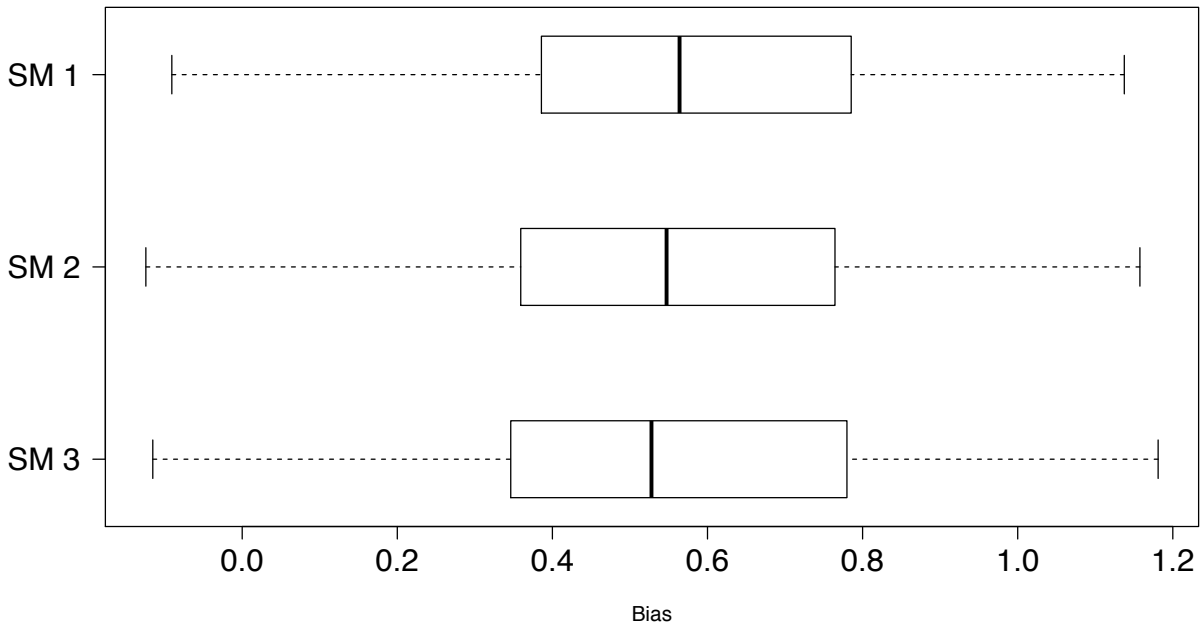


Figure A.2.39- Bias for 28-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories.

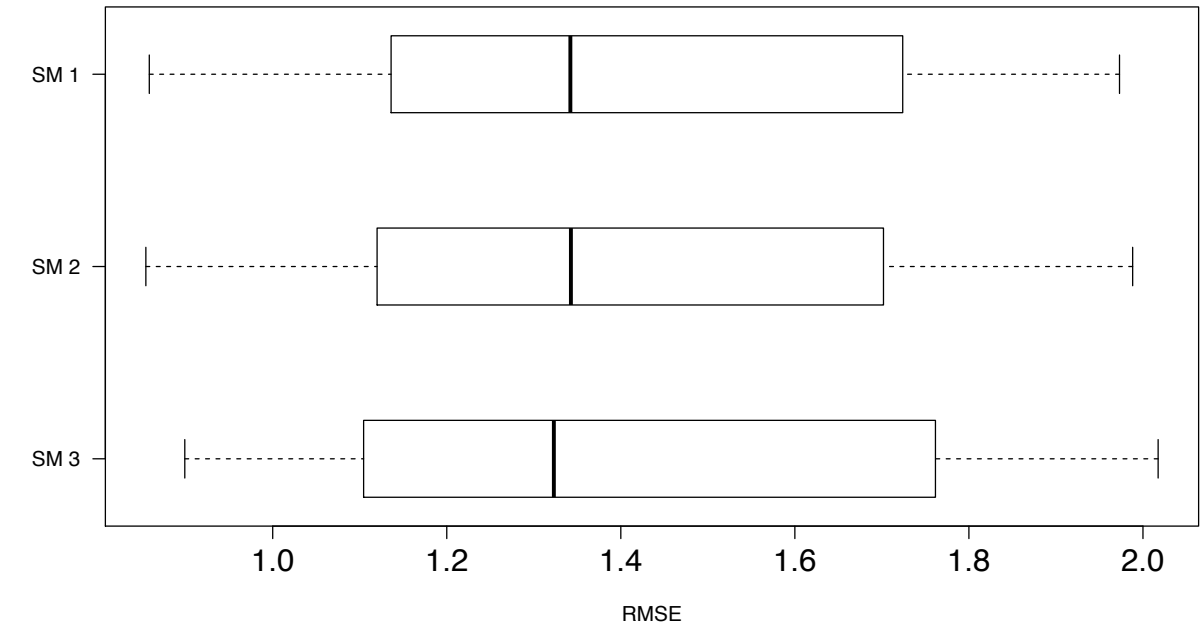


Figure A.2.40- RMSE for 28-day average NWM soil moisture drought categories as compared to county USDM drought weighted average (WA) categories

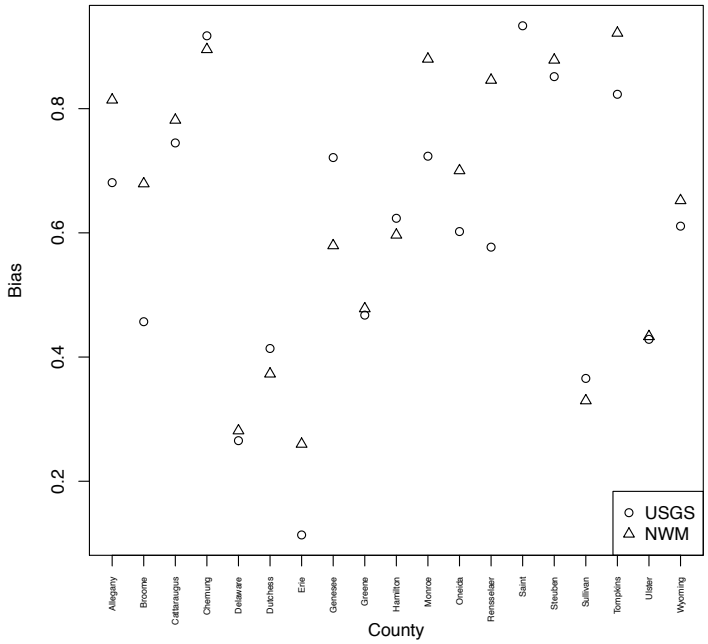


Figure A.2.41- Bias for 28-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

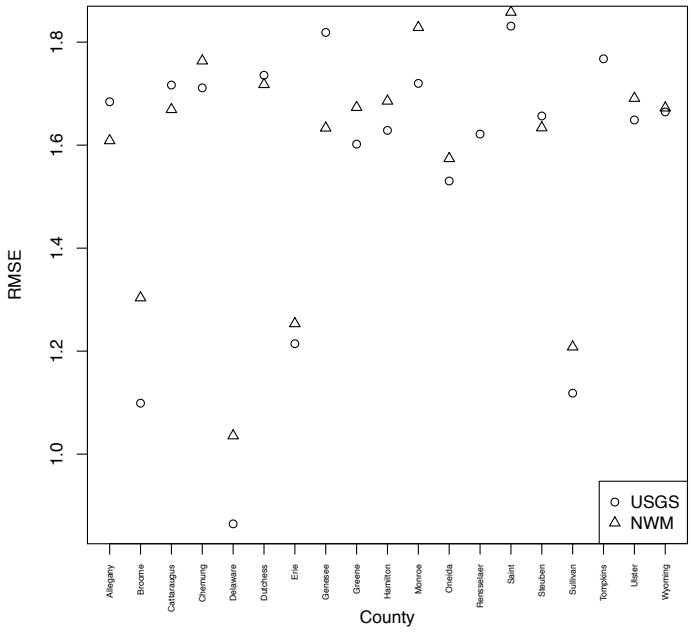


Figure A.2.42- RMSE for 28-day average USGS and NWM streamflow drought categories as compared to county USDM drought majority (M) categories.

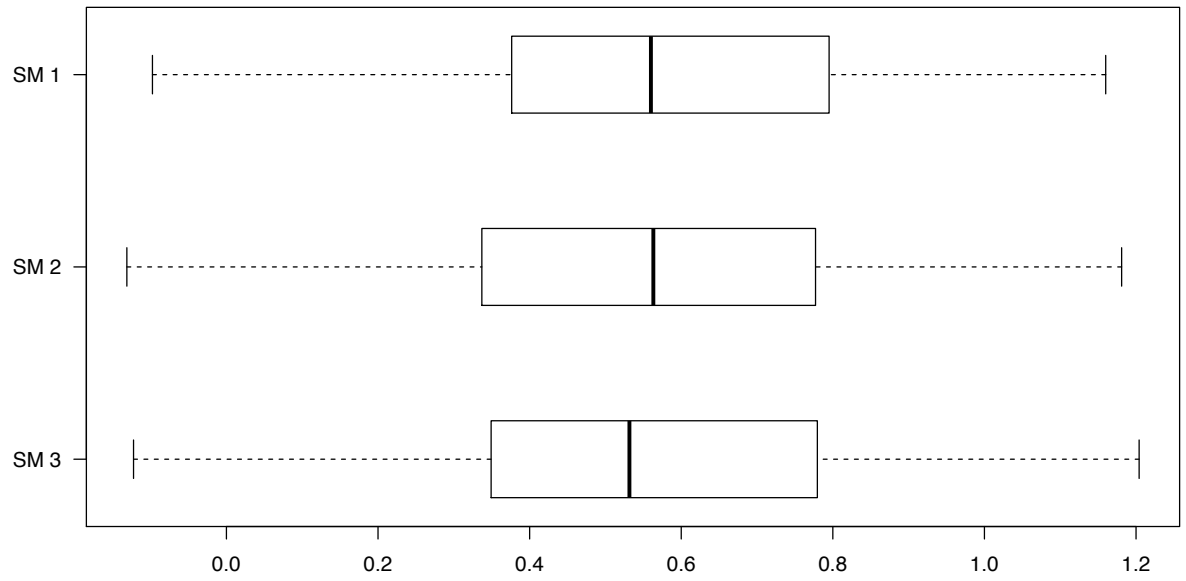


Figure A.2.43- Bias for 28-day average NWM soil moisture drought categories as compared to county USDM drought majority (M) categories.

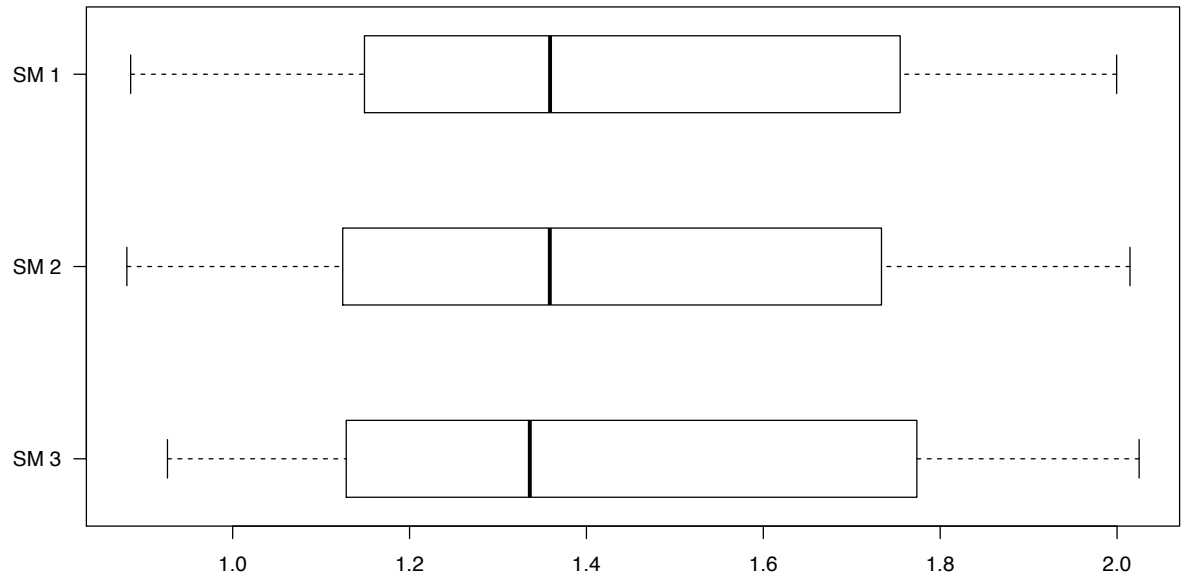


Figure A.2.44- RMSE for 28-day average NWM soil moisture drought categories as compared to county USDM drought majority (M) categories

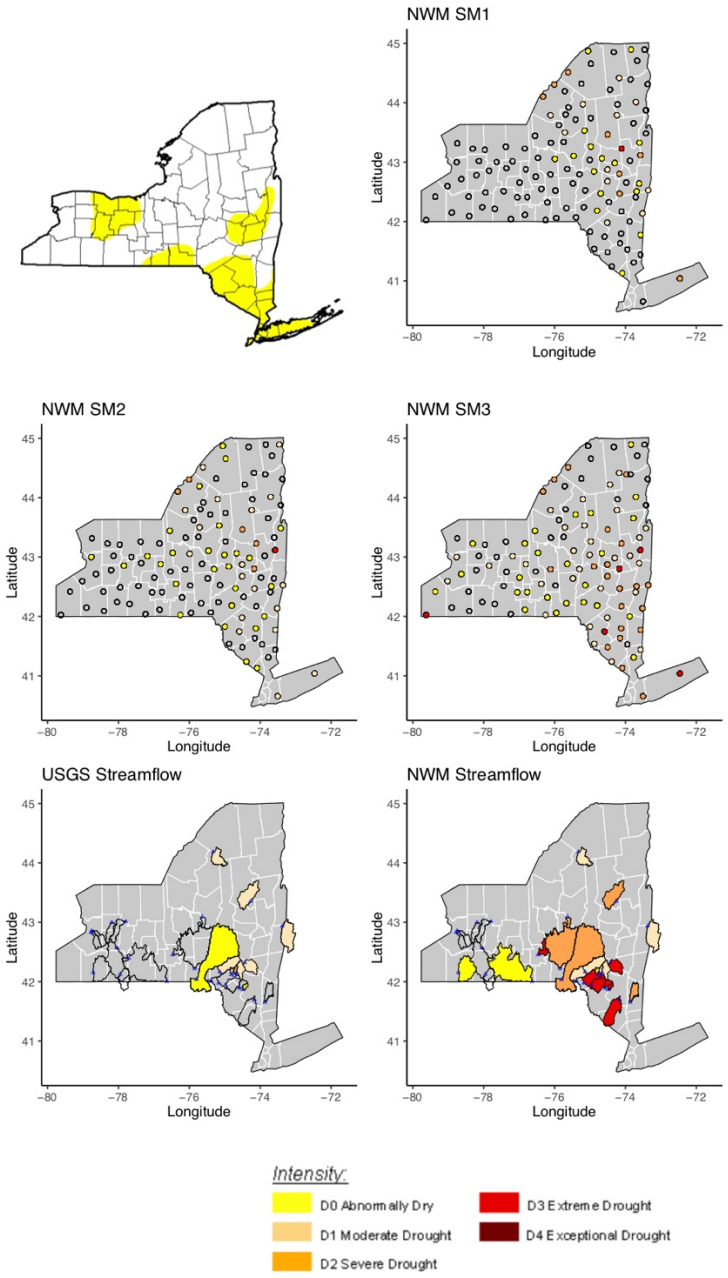


Figure A.2.45- May 3, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

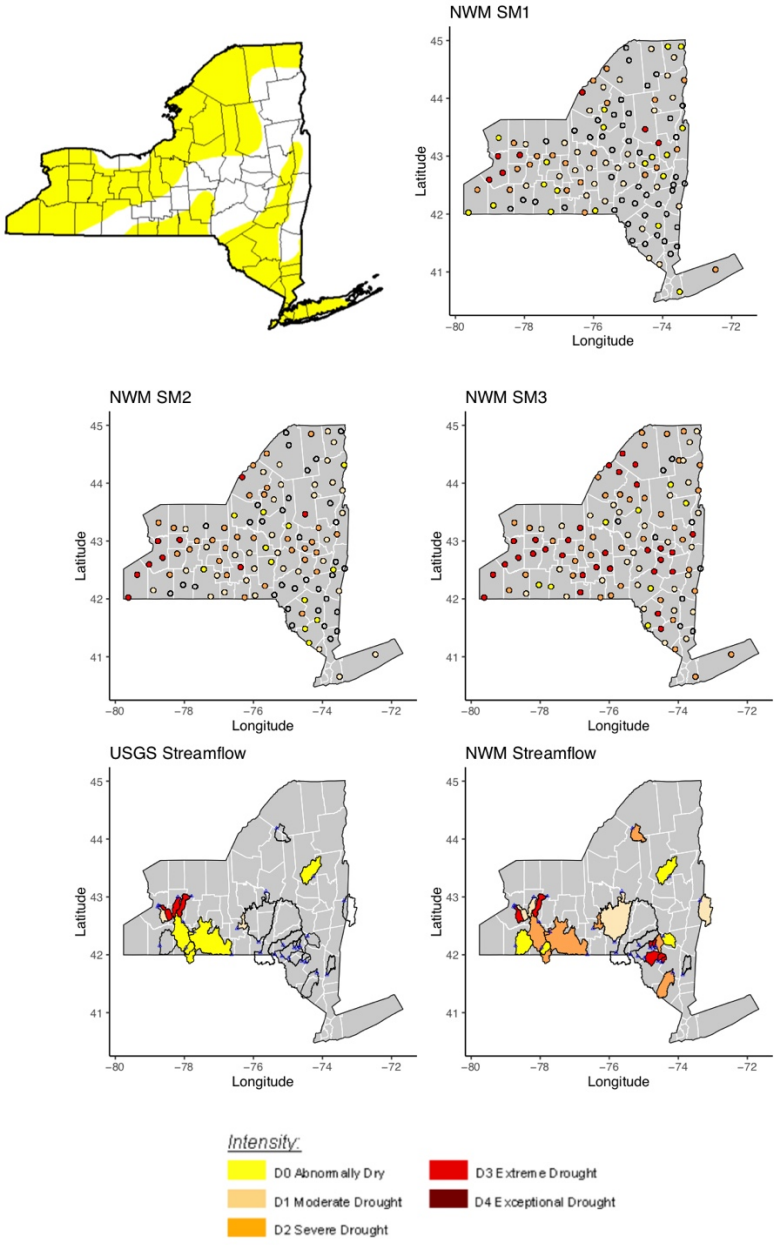


Figure A.2.46- June 7, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

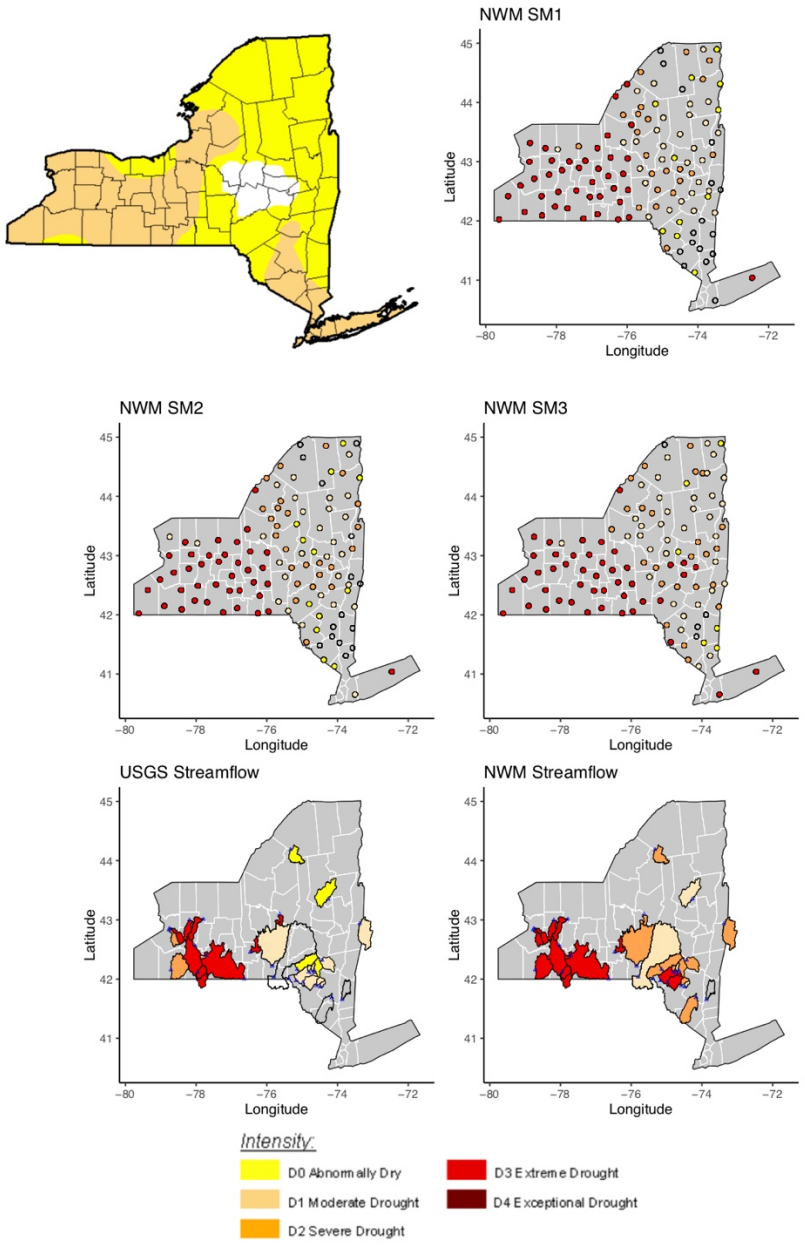


Figure A.2.47- July 5, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

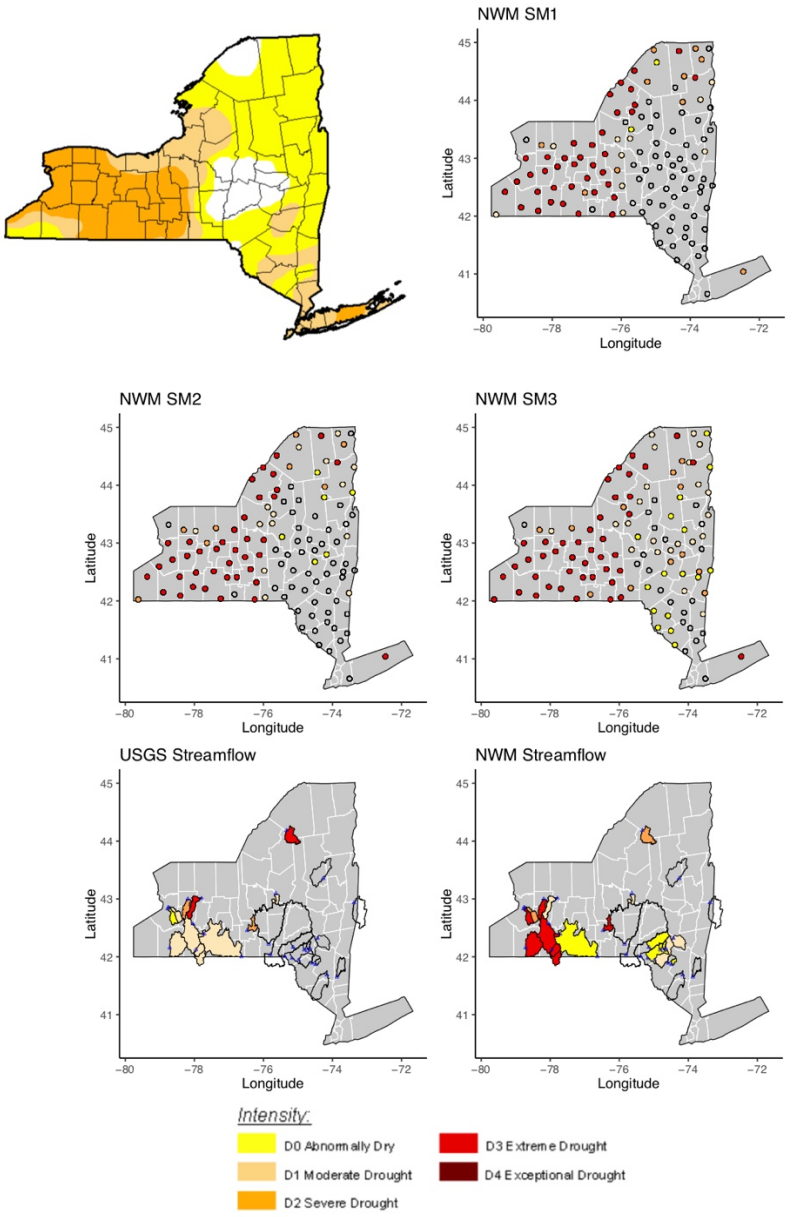


Figure A.2.48- August 2, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

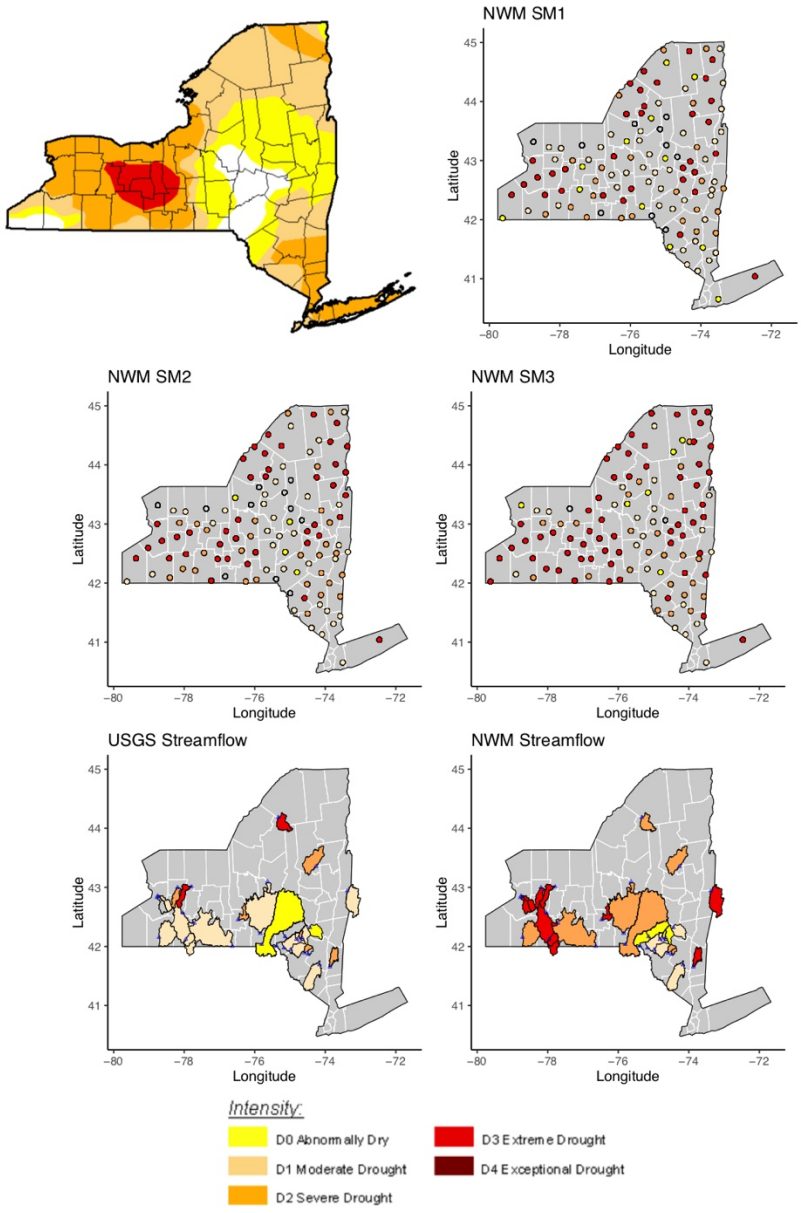


Figure A.2.49- October 4, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

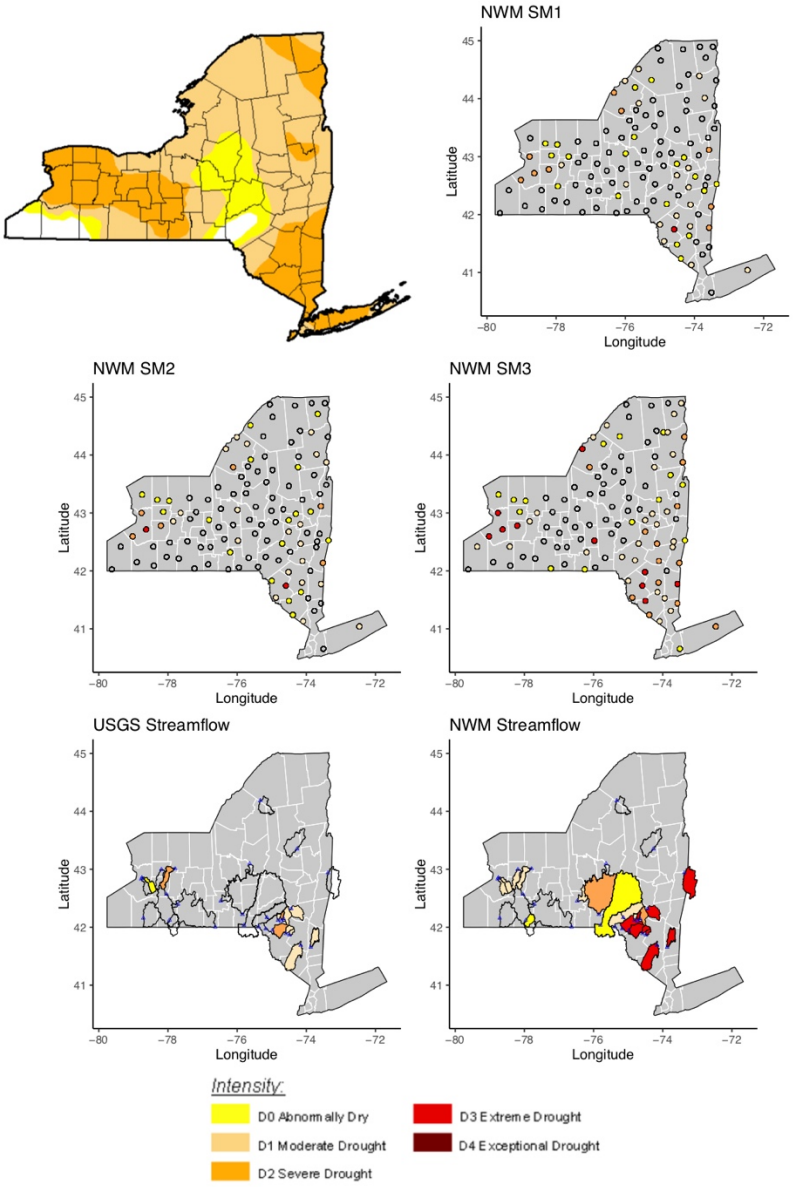


Figure A.2.50- November 1, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

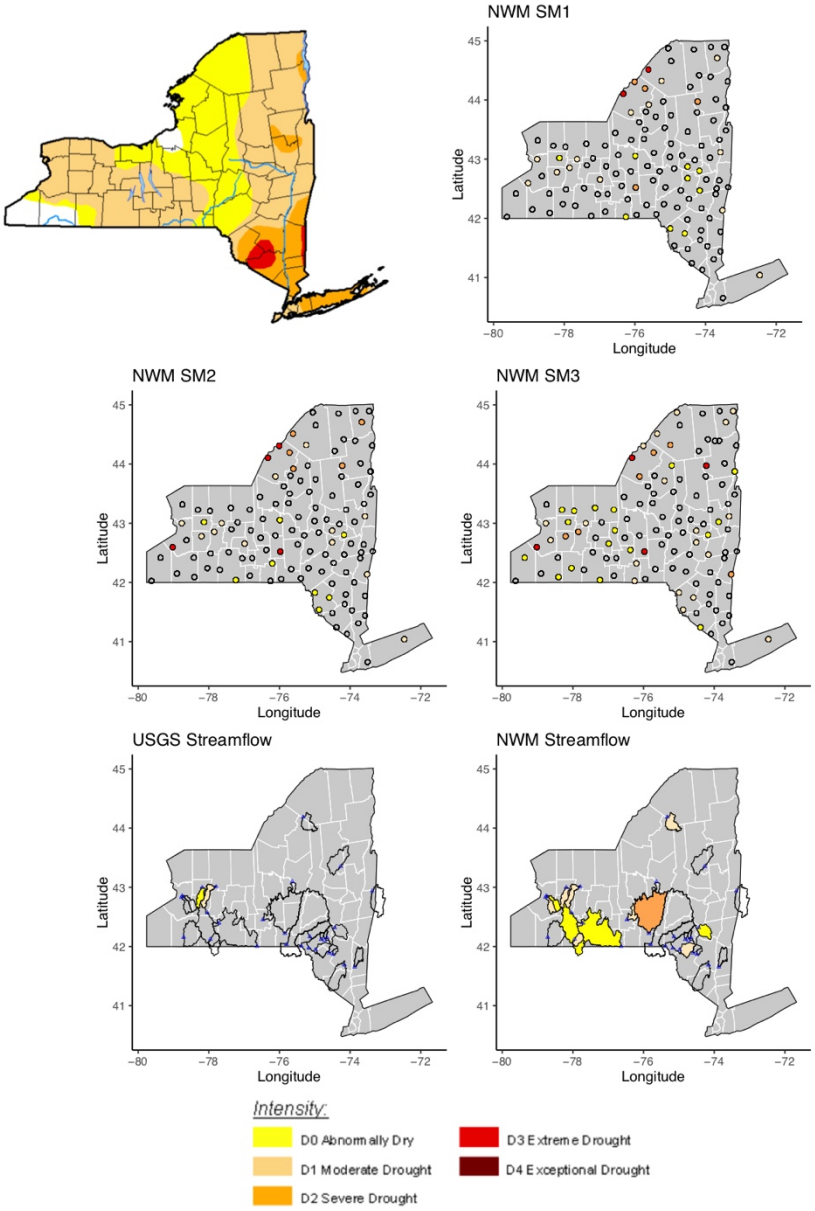


Figure A.2.51- December 6, 2016 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

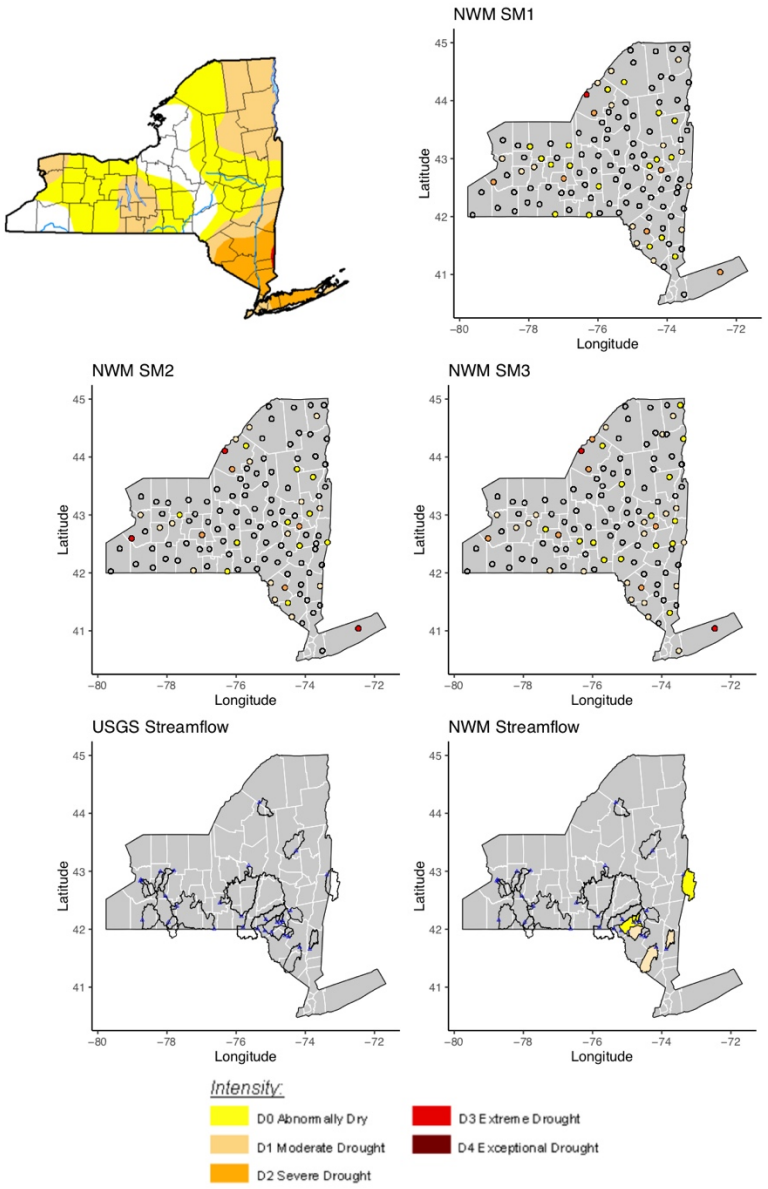


Figure A.2.52- January 3, 2017 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

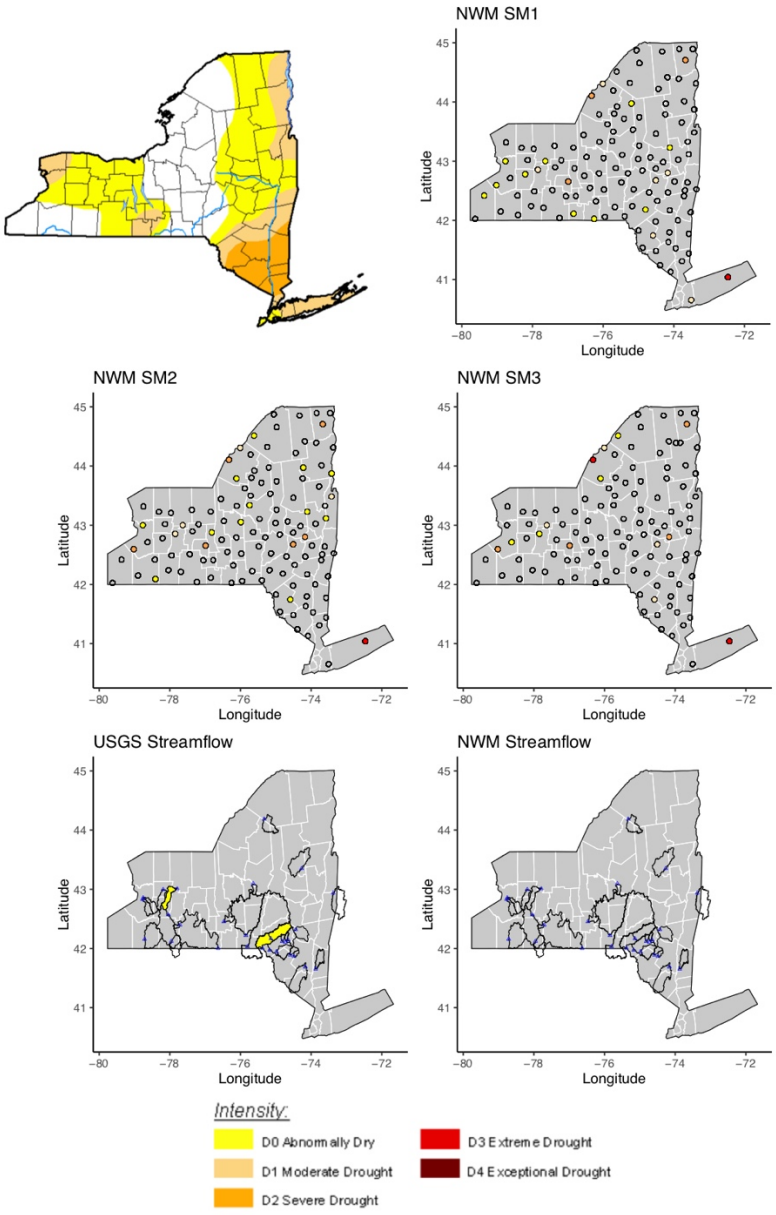


Figure A.2.53- February 7, 2017 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

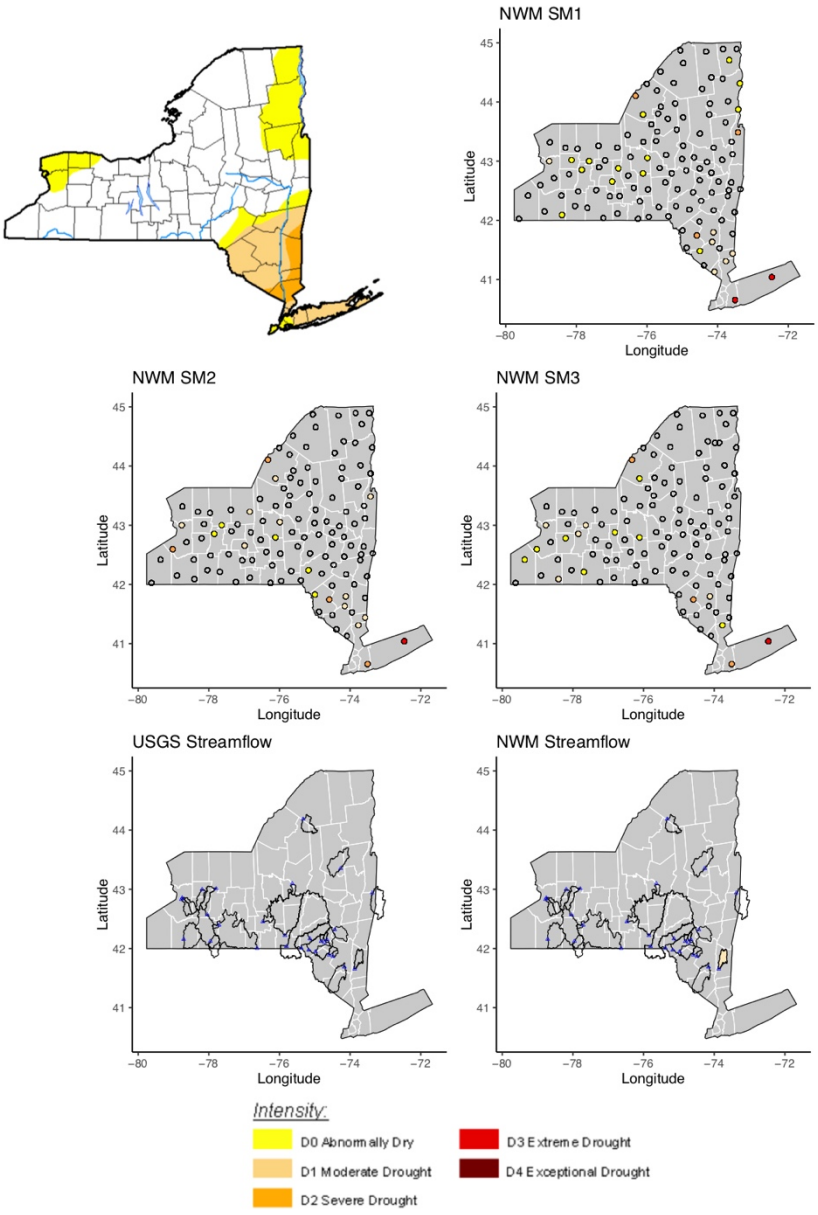


Figure A.2.54- March 7, 2017 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

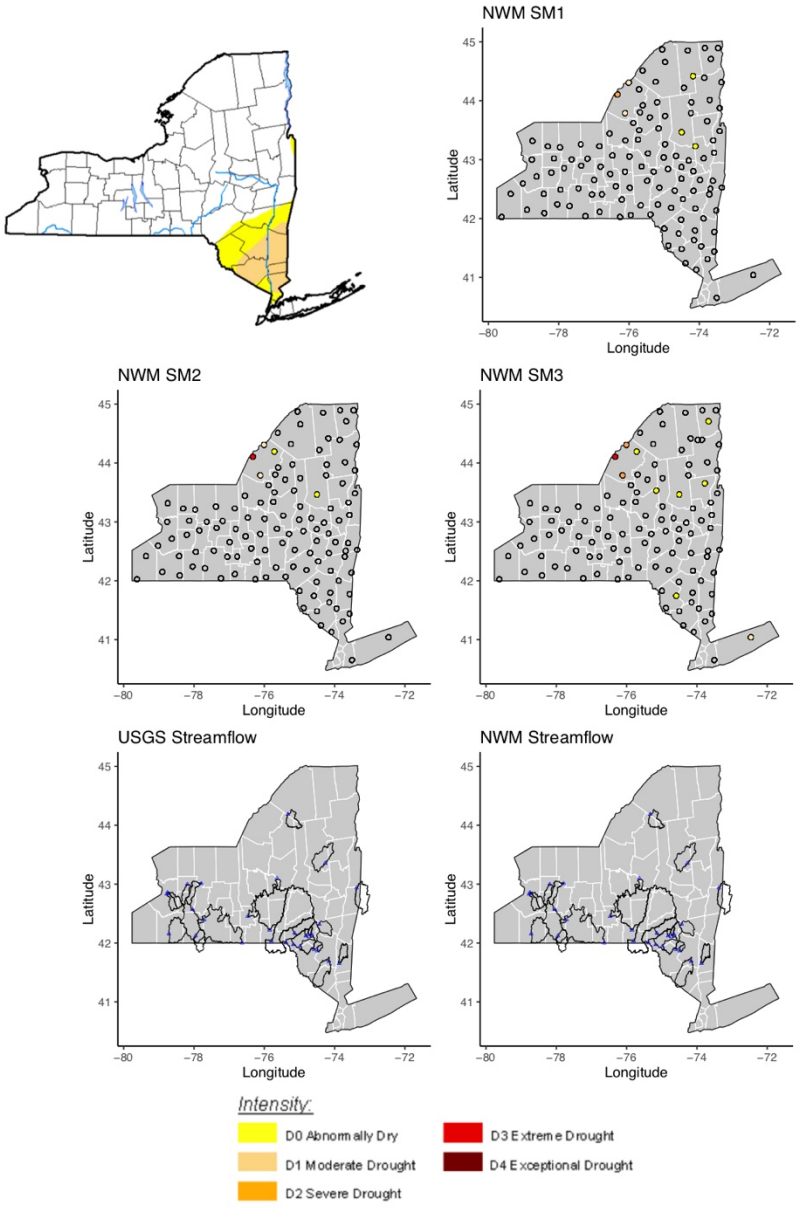


Figure A.2.55- April 4, 2017 drought categories from (a) US Drought Monitor, (b) NWM SM1 profile, (c) NWM SM2 profile, (d) NWM SM3 profile, (e) NWM streamflow, and (f) USGS streamflow.

Vitae

Name: Brenden H. Covert

Date and Place of Birth: October 26, 1994.
Johnson City, New York

Education

Name and Location	Dates	Degree
Seton Catholic Central High School Binghamton New York	Sept 2009 - June 2013	H.S. Diploma
SUNY College of Environmental Sciences and Forestry Syracuse New York	Aug 2013 - May 2017	B.S.
SUNY College of Environmental Sciences and Forestry Syracuse New York	Aug 2017 - Aug 2019	M.S.

Employment

Company	Dates	Position
Belknap Lumber Inc.	Jun 2013 - Aug 2013 May 2014-Aug 2014	Yard Attendant, Driver
Binghamton Materials Handling	May 2015-Aug 2015	Warehouse Attendant
New York State Department of Transportation	May 2016- Aug 2016 May 2017-Aug 2017	Transportation Construction Inspector
Anchor QEA	Aug 2019 - present	Staff Professional 1